Extending Single-Molecule Fluorescence Spectroscopy Methods to Study Protein Conformation and Dynamics of Complex Biological Systems

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Dominik Hänni

von

Toffen BE

Promotionskomitee

Prof. Dr. Benjamin Schuler (Vorsitz)

Prof. Dr. Amedeo Caflisch

Prof. Dr. Peter Hamm

Zürich, 2012

Summary

How proteins behave in the unfolded state and how they sample the available conformational space for their native structure is a subject of ongoing research and dispute. The discovery of molecular chaperones which assist misfolded and unfolded proteins in finding their native structure has added an additional dimension of complexity to these questions. Confinement and molecular crowding as found within a living cell transforms our concept of protein folding as a reaction driven by intramolecular interactions to a process influenced by various intra- and intermolecular forces and effectors. In the work presented here, we systematically investigated fundamental aspects of protein folding, dynamics and protein-protein interactions. This was facilitated by the use of several biophysical techniques, particularly single-molecule Förster resonance energy transfer (FRET). To overcome intrinsic limitations in the classical FRET approach, such as the reduced sensitivity at short distances, novel experimental strategies were developed.

Precise control of the sample temperature on the confocal microscope together with chemical additives to enhance fluorophore stability and brightness allowed us to study the dimension of the unfolded state of the small cold shock protein Csp*Tm* and of the intrinsically disordered protein prothymosin α at various temperatures and denaturant concentrations. For both proteins, which differ significantly in terms of chemical composition and folding behavior, a remarkable compaction at increased temperatures was found. A systematic investigation of the underlying effects together with replica exchange molecular dynamics simulations in explicit water revealed intramolecular hydrogen bonds as the the most probable source for the compaction at higher temperatures.

Different variants of fluorescently labeled bovine liver rhodanese, a highly aggregation prone two-domain protein, were used as a model system for protein-chaperone interactions. Refolding rhodanese in presence of GroEL led to a quantitative binding to the chaperone. We studied the conformations and dynamics of rhodanese in the resulting complexes with GroEL using single-molecule spectroscopy and time-resolved anisotropy experiments. To extend the accessible timescale of the single-molecule experiments, surface immobilization experiments were performed. From these experiments and Brownian dynamics Monte Carlo simulations of orientationally restricted fluorophores we obtained information about the conformation and dynamics of the GroEL bound rhodanese. It was possible to show that rhodanese is bound to GroEL as a partially structured folding intermediate with a rather well-defined ensemble of conformations. Thorough investigation of the inter-dye dynamics from ns to hundreds of ms revealed no observable long-range dynamics within the bound rhodanese. In addition to the biological implications of these experiments such as the lack of long-range dynamics, we also showed how to obtain structural information in presence of restricted fluorescent dye motion.

Next, the folding kinetics of rhodanese within the chaperonin cavity, formed by GroEL and its co-chaperone GroES, were measured using single-molecule FRET experiments. In order to perform time-resolved experiments of the initial binding of GroES to the GroEL rhodanese complex, we combined single-molecule FRET with microfluidic mixing. To acquire more specific information about the enthalpic and entropic contributions of the chaperone-substrate interactions during folding, temperature dependent measurements were carried out. In contrast to the popular picture that suggests an active assistance of folding in addition to the prevention of aggregation, we found a rather unexpected picture for chaperone-assisted rhodanese folding. Whereas the folding rate of the N-terminal domain of rhodanese was not affected by the encapsulation within GroEL/ES, the C-terminal domain showed a significant deceleration in folding. From these results we concluded that there is no universal mechanism of GroEL/ES action. Depending on the substrate and the resultant intra- and intermolecular interactions, the influence on the folding speed and the specific mechanism of GroEL/ES action may be diverse.

Finally, a novel experimental approach that combines FRET and photoinduced electron transfer (PET) quenching and therefore delivers information about short and long-range dynamics within the same molecule at the same time was developed. As a test system, peptides consisting of stiff polyprolines carrying the FRET dyes and a tryptophan residue on a flexible linker as a PET quencher were produced. High resolution fluorescence correlation data obtained from these peptides were fitted with advanced photon statistical models. Additionally, the N-terminal domain of HIV-1 integrase, an intrinsically disordered protein, was studied using this novel approach. The PET quenching of the introduced fluorescent dyes was caused by a single intrinsic tryptophan in the native protein sequence. With a tryptophan-free variant as a control, the feasibility of separating the dynamic contributions of FRET and PET dynamics was proven. Due to the effect of quenching on the apparent transfer efficiency and therefore the distance information obtained by single-molecule FRET experiments, accurate strategies to measure and separate these contributing factors are crucial to obtain reliable results from single-molecule experiments. Since the FRET and PET probes can be separated onto different interacting proteins, future applications of this approach may include combined studies of multi-protein interactions and dynamics.

Zusammenfassung

Wie sich Proteine im entfalteten Zustand verhalten und wie sie den verfügbaren konformationellen Raum nach ihrer nativen Struktur absuchen, ist ein kontroverses Thema, an dem intensiv geforscht wird. Die Entdeckung von molekularen Chaperonen, welche un- und fehlgefaltete Proteine darin unterstützen, ihre native Konformation zu finden, hat die Komplexität dieser Fragestellung um eine zusätzliche Dimension erweitert. Zelluläre Faktoren wie Molecular Crowding und unterschiedlichste Interaktionspartner transformieren unser Konzept der Proteinfaltung von einer Reaktion, welche durch intramolekulare Interaktion getrieben ist, zu einem Prozess, in welchem intra- und intermolekulare Kräfte und Effektoren eine Rolle spielen. In der hier präsentierten Arbeit wurden fundamentale Aspekte der Proteinfaltung, der Proteindynamik und von Protein-Protein-Interaktionen systematisch untersucht. Um dieses Ziel zu erreichen, wurden verschiedenste biophysikalische Methoden, im speziellen Förster Resonanz Energie Transfer (FRET) mit Einzelmolekülauflösung, verwendet. Um intrinsische Beschränkungen des klassischen FRET Ansatzes, wie z.B. die verminderte Sensitivität bei kurzen Distanzen, zu überwinden, wurden neuartige experimentelle Strategien entwickelt.

Die präzise Kontrolle der Probentemperatur auf dem konfokalen Mikroskop zusammen mit chemischen Zusätzen, welche die Stabilität und Helligkeit der verwendeten Fluoreszenzfarbstoffe verbesserten, erlaubte es uns, die Ausdehnung des entfalteten Zustandes für das kleine Kälteschockprotein Csp*Tm* und für das intrinsisch unstrukturierte Protein Prothymosin α bei unterschiedlichen Temperaturen und Konzentrationen an Denaturierungsmitteln zu messen. Für beide Proteine, welche sich in ihrer chemischen Zusammensetzung signifikant unterscheiden, wurde eine bemerkenswerte Kompaktierung der entfalteten Zustände bei erhöhten Temperaturen gefunden. Eine systematische Untersuchung der zugrundeliegenden Effekte, zusammen mit Replica Exchange Simulationen in explizitem Wasser, zeigte, dass Wasserstoffbrücken die wahrscheinlichste Ursache für die Kompaktierung bei höheren Temperaturen sind.

Unterschiedliche Varianten von mit Fluoreszenzfarbstoffen markierter Rhodanese, einem stark aggregationsanfälligen Zweidomänenprotein aus der Rinderleber, wurden als Modellsystem für Protein-Chaperon-Interaktionen verwendet. Rückfaltung von Rhodanese in Gegenwart von GroEL führt zu einer quantitativen Bindung an dieses Chaperon. Wir untersuchten die Konformationen und die Dynamik der gebundenen Rhodanese mittels Einzelmolekülspektroskopie und zeitaufgelösten Anisotropiezerfallsexperimenten. Um die zugängliche Zeitskala der Einzelmolekülexperimente zu erweitern, wurden Immobilisierungsexperimente auf Oberflächen durchgeführt. Von diesen Experimenten, zusammen mit Monte Carlo-Simulationen von in ihrer Orientierung eingeschränkten Fluoreszenzfarbstoffen, war es möglich, Informationen über die Konformation und Dynamik von GroEL-gebundener Rhodanese zu gewinnen. Es war möglich, zu zeigen, dass Rhodanese als partiell strukturiertes Intermediat bestehend aus einem eher gut definierten Ensemble an Konformationen an GroEL gebunden wird. Eine gründliche Untersuchung der Dynamik zwischen den Fluoreszenzfarbstoffen in einem Zeitbereich von Nanosekunden bis zu Zehntelssekunden zeigte keine messbare Dynamik innerhalb der gebundenen Rhodanese. Zusätzlich zu den biologischen Implikationen dieser Experimente, wie dem Fehlen einer langreichweitigen Dynamik, konnten wir zeigen, wie es trotz eingeschränkter Beweglichkeit der Fluoreszenzfarbstoffe möglich ist, strukturelle Informationen zu erhalten.

Im nächsten Schritt untersuchten wir mittels Einzelmolekül-FRET-Experimenten die Faltungskinetik von Rhodanese innerhalb der Kavität, die durch GroEL und das Co-Chaperon GroES gebildet wird. Um zeitaufgelöste Experimente der initialen Bindung von GroES an die GroEL-Rhodanese Komplexe durchführen zu können, kombinierten wir Einzelmolekül-FRET mit Mikrofluidik-Mischexperimenten. Um spezifischere Informationen über die enthalpischen und entropischen Beiträge der Chaperon-Substrat-Interaktionen während der Faltung zu erhalten, wurden temperaturabhängige Experimente durchgeführt. Im Gegensatz zu der klassischen Vorstellung, nach welcher neben der Verhinderung von Aggregation eine akti-

ve Unterstüzung der Faltung angenommen wird, fanden wir ein eher unerwartetes Bild für die Chaperon-assistierte Rhodanese-Faltung. So war die Faltungsrate der N-terminalen Domäne von Rhodanese durch die Enkapsulierung in GroEL/ES unverändert, wohingegen die C-terminale Domäne eine signifikante Verlangsamung in der Faltungsgeschwindigkeit zeigte. Von diesen Resultaten konnten wir schlussfolgern, dass es keinen universellen Mechanismus für die Funktion von GroEL/ES gibt. Abhängig vom Substrat und den daraus resultierenden intra- und intermolekularen Interaktionen kann der spezifische Mechanismus und damit der Einfluss auf die Faltungsrate unterschiedlich sein.

Im letzten Teil wurde ein neuartiger experimenteller Ansatz, welcher FRET mit photoinduziertem Elektronentransfer (PET-Quenching) kombiniert, entwickelt. Dieser Ansatz ist in der Lage, gleichzeitig Informationen über kurz- und langreichweitige Dynamik innerhalb eines einzelnen Moleküls zu liefern. Als Testsystem wurden verschiedene Peptide, bestehend aus einem steifen Polyprolin, welches die Fluoreszenzfarbstoffe trägt, und einem Tryptophan als PET Quencher verbunden durch einen flexiblen Linker, hergestellt. Hochaufgelöste Fluoreszenzkorrelationsdaten wurden mit neu entwickelten Methoden der Photonenstatistik gefittet. Zusätzlich untersuchten wir die N-terminale Domäne der HIV-1 Integrase, einem intrinsisch unstrukturierten Protein, mit dieser neuen Methode. Das PET-Quenching der eingeführten Fluoreszenzfarbstoffe wurde durch ein in der nativen Sequenz vorkommendes, einzelnes Tryptophan verursacht. Eine tryptophanfreie Variante diente als Kontrolle, welche nur FRET Dynamik zeigte. Begründet durch die Tatsache, dass Quenching einen grossen Einfluss auf die apparente Transfereffizienz und damit auf die durch Einzelmolekül-FRET-Messungen gewonnene Distanzinformation haben kann, wird deutlich, wie wichtig präzise Strategien sind, um diese Effekte zu messen und zu separieren. Durch die Möglichkeit, FRET und PET-Sonden auf unterschiedlichen, interagierenden Proteine zu plazieren, könnte diese Vorgehensweise zukünftig dazu benutzt werden, kombinierte Untersuchungen von Multiprotein-Interaktionen und Dynamik durchzuführen.

Publications

- F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, and B. Schuler. (2008). **Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy**. Angewandte Chemie (International ed. in English), 47(33), 6184–6188.

- D. Nettels, S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, and B. Schuler. (2009). **Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins.** Proceedings of the National Academy of Sciences USA, 106(49), 20740–20745.

- H. Hofmann, F. Hillger, S. H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. A. Lipman, and B. Schuler. (2010). **Single-molecule spectroscopy of protein folding in a chaperonin cage.** Proceedings of the National Academy of Sciences USA, 107(26), 11793–11798.

- M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, and H. Grubmüller. (2011). **Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach.** PloS one, 6(5), e19791.

- D. Haenni, F. Zosel, D. Nettels, and B. Schuler. **Direct Combination of Single-Molecule Förster Transfer and Photoinduced Electron Transfer in Polypeptides from a Comprehensive Analysis of Photon Statistics.** Manuscript in preparation.

Contents

1 Introduction & Summarizing Discussion

1.1 Challenges of Measuring Protein Dynamics with Single-Molecule FRET

The pioneering experiments of Ha *et al.* in 1996 in which they used near-field scanning optical microscopy (NSOM) to measure FRET between donor and acceptor fluorophores linked to 10 and 20 base pairs (bp) long double stranded DNA helices proved the general posibility of measuring molecular scales below the diffraction limit at a single-molecule level [30]. They concluded that such an approach should be applicable to measure distances and conformational changes within single biological macromolecules. Three years later, Dahan *et al.* realized an experiment in which they were able to measure single-molecule FRET within similar double labeled DNA constructs diffusing free in solution using a confocal microscope [12]. Shortly afterwards, the same group showed that this approach can be expanded to study subpopulations such as the ones produced by cleavage of DNA or unfolding of DNA hairpins [15]. In the same year, Jia *et al.* presented the first single-molecule FRET measurements of folding dynamics [49]. A folding capable peptide derived from a yeast transcription factor was labeled with donor and acceptor fluorescent dyes. Using a confocal microscope and surface immobilization strategies, they were able to follow folding and unfolding trajectories of single molecules in solution [49, 110]. In the next year Deniz *et al.* measured for the first time a denaturation curve of a of a single-domain protein with subpopulation resolution using single-molecule FRET [16]. They labeled chymotrypsin inhibitor 2, a small, single-domain protein site specifically with donor and acceptor fluorescent dyes. Using a confocal microscope they were able to follow the single-molecule FRET signal of the native and unfolded populations as a function of GdmCl concentration. They found interesting effects such as a limited evidence for a shift in transfer efficiency of the unfolded state upon increase of GdmCl concentration which they attributed to an expansion of the unfolded state upon increase of the denaturant concentration [16]. These successful experiments resulted in an exponential growth in papers dealing with single-molecule methods [70]. In the next few paragraphs, the most important aspects of single-molecule FRET experiments are introduced.

Förster Resonance Energy Transfer Based on earlier work by J. and F. Perrin, Theodor Förster derived a quantum theoretical theory about the transition of excited electron energy between similar molecules in solution by dipole-dipole interaction [23]. For the rate of energy transfer, k_{ET}, Förster derived the following equation:

$$k_{ET}(r) = \frac{1}{\tau}\left(\frac{R_0}{r}\right)^6 \tag{1}$$

where τ represents the intrinsic fluorescence lifetime of the donor dye; R_0 is the Förster distance where 50% of the energy is transferred from the donor to the acceptor; and r is the distance between the donor and the acceptor dye. The strong dependence of the transfer rate on distance makes Förster resonance energy transfer suitable as a spectroscopic ruler [105]. This relationship can also be written in terms of transfer efficiency E:

$$E = \frac{R_0^6}{R_0^6 + r^6} \tag{2}$$

R_0 depends on the overlap integral, J, of the normalized donor fluorescence, $F_D(\lambda)$, and the acceptor absorbance spectrum, $\varepsilon_A(\lambda)$,

$$J = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda, \tag{3}$$

the donor fluorescence quantum yield, Q_D, in absence of an acceptor, the refractive index of the medium between the dyes, n, and Avogadro‘s number, N_A. κ is the orientational factor, which accounts for the relative orientation of the dipole moments of the two fluorescent dyes. κ may take any value between 0 and 4. If the relative reorientation of the two dyes is significantly faster than the fluorescence lifetime of the donor, this effect averages out and κ becomes 2/3. The following equation shows the dependence of R_0^6 on the discussed parameters:

$$R_0^6 = \frac{9000(ln10)\kappa^2 Q_D J}{128\pi^5 n^4 N_A} \tag{4}$$

The spectral and photophysical properties of the most frequently used FRET dye pairs for single-molecule experiments result in a value of R_0 in the range of 35-65 Å [25]. In single-molecule FRET experiments, there are two common ways to calculate the transfer efficiency of a single distance from a fluorescence event: Either it can be calculated using the corrected photon counts for donor $\mathrm{n_D}$ and acceptor $\mathrm{n_A}$ emission

$$E = \frac{n_A}{n_A + n_D} \tag{5}$$

or by measuring the fluorescence lifetime of the donor in presence, τ_{DA}, and absence, τ_{D}, of an acceptor dye

$$E = 1 - \frac{\tau_{DA}}{\tau_D}. \tag{6}$$

These two independently measured transfer efficiencies can either be used to refine the experimental values or to check for anomalies in the experimental setup or the specimen under investigation [60, 33, 93].

Protein Purification and Labeling The protein of interest is typically produced by solid phase peptide synthesis or by recombinant expression combined with site directed mutagenesis [93]. Common strategies to introduce fluorescent dyes are by labeling naturally occurring amino acids such as lysine or the N-terminal amino group with N-hydroxy-succinimidyl ester, or cysteine with maleimide-containing moieties [99, 55]. Examples of fluorescent dyes commonly used for single-molecule FRET studies are Alexa Fluor 488, Cy3, Atto 550 and Alexa Fluor 555 as donor dyes and Alexa Fluor 594, Cy5, ATTO 647N, Alexa Fluor 647 as acceptor dyes [95, 92]. To avoid fluorescent contamination, only the highest quality of buffers and reactants should be used for all the reactions and purifications by high-performance liquid chromatography (HPLC) or fast protein liquid chromatography (FPLC) [93].

Measuring Single Molecules in Diffusion A schematic overview of the instrumentation to measure fluorescent single molecules in solution is shown in figure 1. A pulsed or continuous wave laser is reflected by a dichroic mirror into a high numerical aperture oil or water immersion objective. The objective focuses the laser light to a diffraction-limited spot a few micrometers into the sample. The picomolar concentration of the sample results in the excitation of a single molecule at a time within this excitation volume. The emitted fluorescence light is collected by the same objective and filtered for backscattered laser light by the dichroic mirror. The fluorescence light is focused onto a pinhole to eliminate out of focus light and therefore to improve the axial resolution and reduce the background. This confocal setup, which increases the spacial resolution to the diffraction limit, is a prerequisite for doing single-molecule fluorescence experiments on freely diffusing molecules in solution. To aquire information about the fluorescence anisotropy, the next element is a polarizing beam splitter that transmits parallel polarized photons and reflects perpendicular ones. After the polarizing beamsplitter, the donor and acceptor photons are separated by dichroidic beam splitters and focused on single photon avalanche photodiodes (SPADs). These cooled semiconductor photodetectors can have up to 60% quantum efficiency in the visible spectrum. Within the time resolution of around 350 ps, these detectors will convert each detected photon into a digital output pulse [71, 93, 94]. Counting electronics with ps resolution assign a time stamp to each pulse of the detectors [120]. This photon arrival time information for each detector is collected and stored for subsequent analysis.

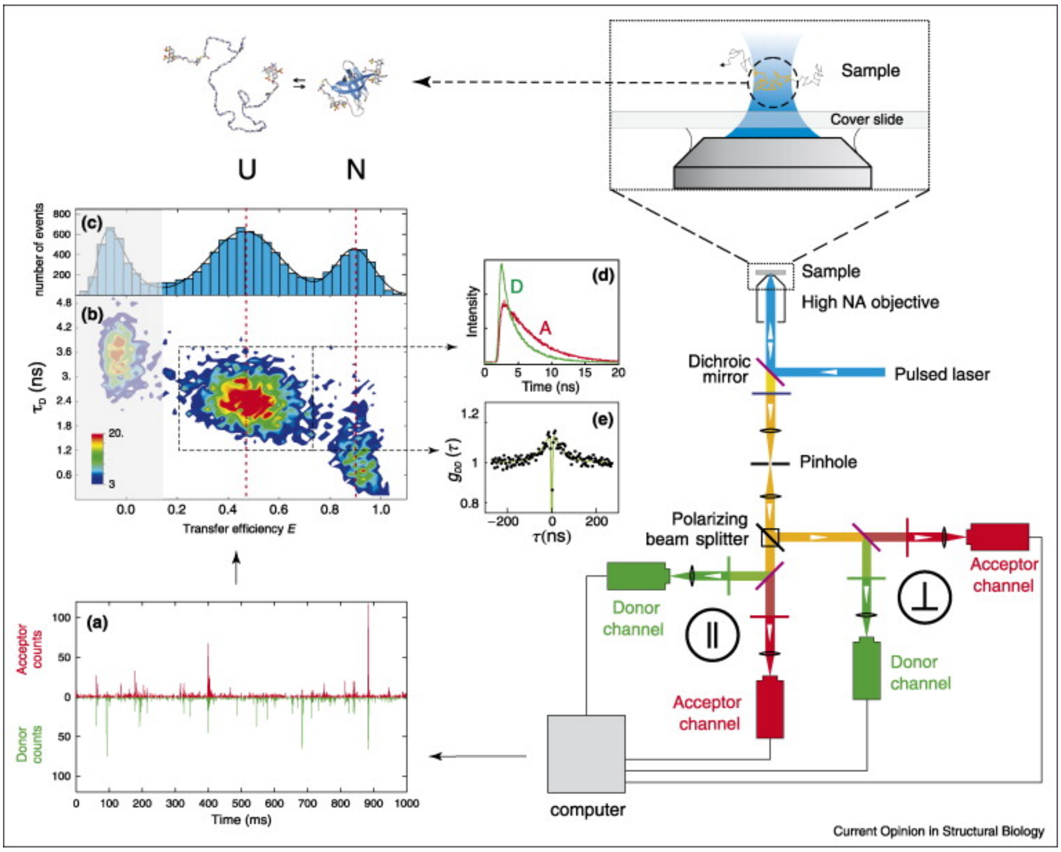


Figure 1: Schematic overview of confocal single-molecule fluorescence data acquisition and analysis. On the right hand side, the most important components of of a four channel confocal microscope are shown. **(a)** Signal trajectory calculated from the photon arrival data collected by photon counting hardware and a computer. **(b)** Pulsed laser excitation allows to calculate transfer efficiency and lifetime of every single-molecule event. **(c)** Histogram of collected transfer efficiencies can reveal different FRET populations such as native N and unfolded U state populations **(d)** Subpopulation-specific fluorescent lifetimes for donor and acceptor fluorescent dye. **(e)** FCS measurements at single-molecule concentration allow calculation of subpopulation-specific correlation functions and therefore to obtain information about chain reconfiguration times (Adapted from Schuler 2008 [94])

Data Analysis A time binning of the photon arrival data can be used to calculate a signal trajectory for donor and acceptor photons as presented in figure 1(a). In the simplest approach, the sum of the donor and acceptor photons within a time-bin is taken and fluorescence bursts are selected based on a threshold criterion. After applying corrections for differences in quantum yields of the fluorophores, fluorescence background, and crosstalk between the donor and the acceptor channels, the transfer efficiency is calculated according to equation 5. These transfer efficiencies can be collected into a transfer efficiency histogram as shown in figure 1(c) [17, 93]. In addition to this simple but informative analysis, there are plenty of alternative types of information to extract from the single-molecule data. By using a pulsed laser, it is possible to record the arrival time of photons relative to the laser pulse, which allows to extract fluorescence lifetime data from each individual single molecule as shown in figure 1(d) [21]. Since the fluorescence can be sorted for polarization, as shown in figure 1, fluorescence anisotropies can be calculated for each burst.

In FCS analysis, the time-traces of photon arrival times or the inter-photon-times can be correlated. From the fluctuation in fluorescence intensity over time, information about translational diffusion constants and the number of molecules in the confocal volume can be derived [57]. By experimentally overcoming the limitations in time resolution of the SPADs and by using state of the art photon counting hardware, it is possible to expand the accessible timescale of such experiments into the ns-range and therefore to observe photophysical effects such as photon antibunching and fast, diffusive chain dynamics of unfolded proteins [74].

1.1.1 Confinement and Rotational Effects

Time dependent fluorescence experiments with polarized excitation and detection schemes can give insights about temporal variations in dipole orientation of fluorescent probes. Within the context of fluorescently labeled macromolecules, this information can lead to an understanding of the rotational motion of the whole macromolecule or of individual subunits [122]. By using intrinsic fluorescence reporters such as the naturally occurring amino acid tryptophan, its even possible to measure protein flexibility and dynamics without further chemical modification [73]. However, Szabo enumerated several complications found in non-ideal experimental systems: the overall motion can be anisotropic; restricted internal motion can exist; the excited state can decay to the ground state; energy transfer may occur; the probe can be trapped in conformations with different emission characteristics, and there may be several excitable and interconverting electronic states [107].

If the relative orientations of the donor and the acceptor dye are randomized during the excited state lifetime, the orientation factor κ^2 becomes 2/3 [106]. In the case in which the orientation of the dyes is fixed, but the relative orientation is randomly distributed, the value will be 0.476 [103, 104]. For the situations in between these two extremes, Lipari and Szabo proposed to address this restricted motion by a "wobbling in the cone" model introduced by Kawato *et al.* to describe fluorescence data of molecular motions in liposomes [53, 65]. In our work on GroEL-bound Rhodanese, we found that the binding of the fluorescently labeled rhodanese to the GroEL chaperone results in a decrease of the rotational freedom of the fluorescent dyes (See section 1.2.3). Using Monte Carlo simulations of fixed but randomly distributed donor and acceptor cones and a normal distribution of inter-dye distances, it was possible to reproduce the ensemble anisotropy decay experiments and the single-molecule transfer efficiency histograms. This approach allowed us to retrieve information about the distance distribution even under conditions of highly restricted dye motion [35].

In 1976, Ehrenberg and Rigler proposed an experimental strategy to measure the rotational diffusion of macromolecules using fluorescence correlation spectroscopy as an alternative to light scattering [22]. Almost 10 years later, Kask *et al.* showed the possibility to measure rotational correlation times of rhodamine labeled bovine carbonic anhydrase B using high resolution FCS experiments (ns-FCS) [50]. Thereafter, the same group proposed a methodology to separate rotational effects from other effects contributing to the fast dynamics found in such ns-FCS experiments. By having a perpendicular orientation between the excitation laser and the collection system for the emitted fluorescence light, they were able to measure correlations for all possible combinations of excitation and emission polarizations [51]. Using an epifluorescence microscope setup, as illustrated in figure 1, has the disadvantage that the excitation of the fluorescent dye and the collection of the emitted fluorescence light is done trough the same, high numerical aperture objective. Therefore, there is no perpendicular orientation between fluorescence excitation and detection, and a complete separation of the different polarizations is not possible. An approach such as that proposed by Kask *et al.* to separate rotational effects from other effects such as those introduced by the fast dynamics between the FRET fluorophores, is very difficult to implement [50]. Consequently, unrestricted movement of the attached fluorescent dyes is an obligatory requirement for measuring the chain dynamics using FRET ns-FCS.

1.1.2 Accessible Timescales and Fluorophore Stability

In measuring the dynamics of freely diffusing molecules with single-molecule FRET using FCS, there are two natural time limitations. The first is due to the inherent lifetime of the excited state of the fluorescent molecule, which sets the lower observation limit for dynamic processes. The second factor is that the investigated molecules need a limited time to diffuse trough the confocal volume. Depending on various parameters, this timescale is typically in the range of 0.1-2 ms and therefore sets the upper time limit for the observation of a single molecule.

To expand the accessible timescales of such single-molecule experiments, various strategies were introduced and applied over the last few years. Surface immobilization techniques to fix the macromolecules of interest onto functionalized surfaces directly or encapsulated in tethered lipid vesicles allows observation of single molecules for much longer timescales [31, 90]. In fact, the only limiting aspect of such experiments is the irreversible photo-destruction of the involved fluorophores [92]. An alternative strategy to get information about dynamics longer than the diffusion limit emerged from the observation that a freely diffusing molecule has a certain probability to pass the confocal volume multiple consecutive times which led to the recent development of the recurrence analysis of single particles (RASP) [40]. RASP allows one to measure interconversion kinetics of different FRET populations for longer timescales than imposed by the diffusion time of a single molecule trough the confocal volume. To get information about ns timescales, such as the reconfiguration time of an unfolded protein chain, novel FRET FCS approaches were implemented by Nettels *et al.* and used in this work [74]. By accumulating a large amount of single-molecule data, it is possible to get interphoton-times in the time-range of these effects. By correlating such datasets, it is possible to determine the underlying rate constants [74].

In order to investigate the chain dynamics of rhodanese bound to and encapsulated within GroEL, several developements, outlined fully elsewhere in this work, were required. This was achieved by optimizing existing experimental and theoretical treatments in order to aquire the best possible data and extract rigorously tested experimental parameters and by the developement of new software tools for automated data aquisition and analysis (Additional information can be found in the appendix, see section 8). Additionally, a large, automated screen of photo-protection and photo-enhancing agents was carried out. The resulting combinations of chemicals allowed us to increase the brightness of the fluorophores and to reduce photobleaching. Finally, an immobilization strategy was developed, using a non-covalent approach based on PLL-PEG [35, 41].

1.1.3 Quenching of Fluorophores by Natural Amino Acids

As a consequence of observing the quenching of anthracene by diethylanaline, and the resulting anomalous fluorescence signal, Weller formulated a theoretical framework to describe complex formation and electron transfer from a donor to an acceptor molecule [123]. The finding that this photoinduced electron transfer (PET) plays the central role in the photosynthesis of plants and bacteria resulted in a quest for understanding of the involved effects and in attempts to mimic this light harvesting mechanism with artificial compounds [121]. In PET, an electron of the donor molecule, such as a polyaromatic fluorescent molecule, is excited and transferred to an acceptor molecule. Such an electron transfer needs direct interaction of the electron orbitals between the donor and the acceptor molecule and therefore complex formation before or during the excited state lifetime of the donor molecule. Complexes formed during the excited state are called exiplexes. Due to the delocalization of an unpaired electron in such an aromatic system, these radical ions are relatively stable and have a rather low chemical reactivity [2]. This can lead to thermal back-electron transfer as a self-repairing mechanism, avoiding a potential damaging PET process [14]. Since such electron transfer reactions are faster than the fluorescence lifetime of the fluorophores used, PET is the most probable de-excitation process upon complex formation in the presence of a suitable electron acceptor. When the donor molecule is a fluorophore, PET leads to a loss of fluorescence emission and therefore can be used to monitor complex formations in an on/off manner [14]. Natural amino acids such as tryptophan have been shown to be suitable PET acceptors for fluorescent molecules such as rhodamine, oxazine and fluorescein dyes [79, 20]. The strong quenching upon complex formation with tryptophan can be used to measure contact formation rates within peptides and proteins using single-molecule fluorescence spectroscopy. In contrast to FRET, PET-Quenching needs a direct contact between donor and acceptor. This qualifies PET-quenching as a reporter for chain dynamics of unfolded peptides with a fluorophore-Trp distance of 5-30 amino acids [78].

Quenching of fluorescent dyes by natural amino acids is not a characteristic of only some exotic fluorescent dyes, but rather common behavior of most organic fluorophores. This includes the popular Alexa 488/555 and Alexa 488/594 FRET pairs often used to study protein conformations and dynamics on a single-molecule level, which were shown to be quenched by a combination of dynamic and static quenching processes by various naturally occurring amino acids [10]. Such quenching processes can have a profound impact on the experimental results obtained using single-molecule FRET. Depending on the fluorescent dyes and the quenching mechanism, transfer efficiencies and therefore the distance information can be significantly changed to higher or lower values. Importantly, PET quenching leads to an on/off behavior of the fluorophore which contributes to the observed decays in FCS measurements. Since the timescales of PET quenching and protein chain dynamics measured with FRET can be comparable, such additional quenching contributions can have a very significant impact on the results obtained. Most of the second part of the work presented here is dedicated to the question of how quenchers affect single-molecule FRET experiments. Experimental systems and analytical models were developed to unveil these effects and to show the extent of their impact on single-molecule experiments. Since FRET and PET are complementary in terms of distance information, we developed an approach to unify the the two methods to aquire more refined insights into protein dynamics by observing the long-range FRET and short-range PET dynamics in the same molecule at the same time.

1.2 Effect of Intra- and Intermolecular Interactions on Protein Conformation and Dynamics

The pioneering experiments by Anfinsen *et al.* in 1961, in which the unfolding of bovine pancreatic ribonuclease was found to be reversible, lead to the postulation of the so called "thermodynamic hypothesis". In this hypothesis, the three-dimensional structure of native proteins under native conditions is the one with the lowest Gibbs free energy [4, 3]. This work led to the question of how proteins find their native structure. Following computer simulations of folding pathways using a lattice heteropolymer model, Levinthal found that the crucial factor discriminating among topological pathways is the heterogeneity of native contact energies [62]. This finding, together with the fact that proteins can fold rapidly, led him to the conclusion that protein folding is not a random search over all possible chain conformations, but that there is a folding pathway, where local amino acid sequences form stable interactions and therefore act as nucleation points within the folding process [63]. This finding triggered the search for protein folding pathways. The first spectroscopic studies of protein folding revealed single and multi-exponential kinetics, which lead to the postulation of on- and off-pathway folding intermediates [46, 115]. From there on, various attempts to formulate a theoretical framework of protein folding were made. Terms such as “energy landscapes”, in which the free energy of each conformation is a function of the degrees of freedom of the chain, or "folding funnels", in which the vertical funnel axis represents the "effective free energy" of a given conformation which converges towards a minimum or the native state, have been proposed [18]. The hallmark of these more recent approaches to describe protein folding is the assumption that there is not a single protein folding pathway, but the possible interactions within the protein chain of the unfolded state ensemble represent an energy landscape that directs the individual search of each folding protein toward its energy minimum and therefore the native state. In contrast to ensemble experiments, single-molecule experiments have the advantage to offer in the best case direct experimental information about the folding pathway of individual molecules [94]. Using denaturants such as guanidinium hydrochloride (GdmCl) or urea allows the study of the conformation and the dynamics of the unfolded state of proteins. Since each molecule is observed individually, conditions where folded and unfolded molecules are in equilibrium can be used, and the unfolded state and folding dynamics can be studied under almost native conditions [74]. Varying the experimental conditions such as denaturants, temperature, ionic strength, and/or viscosity can give fundamental insights about the underlying driving forces of protein folding and therefore about the shape of the underlying energy landscape [77, 72, 76].

1.2.1 Insights from the Unfolded State of Proteins

To get insights into protein dynamics and protein folding pathways, the unfolded state or to be more precise the denatured state ensemble (DSE) is the logical starting point [102]. Observables such as the radius of gyration (R_g), chain reconfiguration times and the secondary and tertiary structure as assessed by CD-signal as a function of the accessible experimental parameters such as pH, denaturant concentrations and temperature can contribute to a better understanding of the DSE. Additionally there is an increasing interest in the properties of natively unfolded proteins which have little or no ordered structure under physiological conditions [116]. During the last years there was a remarkable increase in publications dealing with these so-called intrinsically disordered proteins (IDPs). It has been shown that a significant number of these proteins fold upon binding to target molecules and are often closely associated with cellular control mechanisms and signaling [126]. This shows that understanding the DSE is not only of crucial importance to the classical protein folding picture, in which proteins cycle between the native and unfolded states, but also has a profound impact on our knowledge of communication and regulation networks within the cell.

In contrast to ensemble methods such as fluorescence spectroscopy and small angle X-ray scattering (SAXS), single-molecule Förster resonance energy transfer (FRET) spectroscopy has the advantage of separating folded and unfolded subpopulations. This allows the study of unfolded state properties under nearly native conditions in which a significant folded population is present. Obviously, these conditions are the ones with the highest physiological relevance. The requirement of working at pico-molar concentrations in single-molecule experiments significantly reduces the problems with protein aggregation [76], a significant advantage when working with unfolded proteins under near native conditions.

1.2.2 Temperature-Induced Collapse of Unfolded Proteins

An important parameter in understanding the DSE is temperature. In contrast to parameters such as denaturant concentration for which independent experiments have shown a continuous expansion of the unfolded state upon addition of denaturant [59, 113, 39], the behavior of the DSE upon a change in temperature is much less clear. SAXS experiments at different temperatures on reduced Ribonuclease A and β-Lactoglobulin showed no change in the R_g of the DSE [47, 52]. Dynamic light scattering (DLS) experiments on reduced Ribonuclease A and Ribonuclease T1 revealed a collapse with increasing temperature [82, 24]. It is important to mention that there is a significant disagreement in measured unfolded state behaviour between single-molecule FRET and SAXS experiments. Various proteins and peptides show a compaction of the DSE upon reduction of the denaturant concentration in single-molecule FRET experiments [98, 69, 39, 102]. In contrast the trend is much less obvious when measured by SAXS [85, 48, 102].

In order to use single-molecule FRET methods to measure the temperature dependent behavior of the DSE, we had to overcome the following experimental complications. Firstly, it was necessary to develop a device which allows accurate temperature control of the sample without interfering with the sensitive confocal optics, such as microscope objectives. Secondly, it was required to screen for combinations of photo-protective agents which decrease photo-blinking and photo-bleaching of the fluorescent dyes that occurs at higher temperatures. Solving these issues allowed us to measure the dimension of the DSE of the small cold shock protein Csp *Tm* and the IDP Prothymosin α for temperatures between 282 and 339 K. By measuring the two proteins over various temperatures and GdmCl concentrations, a significant compaction of the unfolded state with increasing temperature was found. Additional temperature dependent DLS experiments on Csp *Tm* were able to confirm this finding at high GdmCl concentrations. Following the change in the CD-signal at 220 nm over the investigated temperature range showed more than a two-fold increase in signal amplitude for a destabilized variant of Csp *Tm,* which is an indication for the formation of secondary structure at higher temperatures.

As a first attempt to explain these experimental findings we compared the collapse to the temperature dependent binding of GdmCl to the protein chain. Since earlier experiments found a decrease in the binding constant of GdmCl to the protein with increasing temperature [67], one would expect a collapse due to reduced denaturant-protein interactions at higher temperatures. Such a behavior can indeed be found in our measurements. However, extrapolation of the R_g at different temperatures towards zero molar GdmCl reveals that there is no convergence of the R_g for the different temperatures, which implies that there is a temperature-induced compaction even at zero molar GdmCl. Obviously, the dissociation of GdmCl can not be the sole explanation for the compaction at higher temperatures. The hydrophobic effect may be another candidate for compaction of the unfolded state at increasing temperatures. The experimentally found temperature dependence for Csp *Tm* does not resemble the one expected for the hydrophobic effect. To further investigate this effect, similar temperature dependent experiments were performed on Prothymosin α, an IDP with a large proportion of charged and polar amino acids and therefore extremely low hydrophobicity. Within an error of 3 % the same temperature dependent collapse as for Csp *Tm* has been found. This is further evidence against the hydrophobic effect as the main source for the collapse at higher temperatures.

Replica exchange molecular dynamics simulations of unfolded Csp *Tm* in explicit water at different temperatures showed, for certain force fields, a correlation between collapse and intramolecular hydrogen bonding. An increase in temperature increased the number of protein-protein interactions and decreased protein-water hydrogen bonds. However, no clear types of emerging secondary structures at increasing temperatures could be assigned from the available simulated data.

Taken together these experiments show that there is not a single effect contributing to the temperature induced collapse but a variety of effects ranging from hydrophobic effect to seccondary structure formation. Overcoming the technical limitations for single-molecule FRET experiments at increased temperatures allowed us to separate the different contributing effects. From this we were able to conclude that the increase of intramolecular interactions at higher temperatures compensates for the conformational entropy introduced by the increase in thermal energy.

Such insights into the temperature dependence of the DSE can be used to further improve theoretical models and especially force fields used for protein simulations. Best and Mittal expanded the Amber ff03* force field to simulate the unfolded state of proteins at increased temperatures [6]. They used a TIP4P/2005 water model together with the Amber ff03w force field where they reparametrized the backbone dihedral potential corrections compared to ff03*. Simulations on the 15 residue peptide Ac-(AAQAA)$_3$-NH$_2$ showed that the experimental found R_g was reproduced adequately by the simulations. Moreover, with increasing temperature a collapse and an increase in the number of peptide-peptide hydrogen bonds has been found. Such improvements will lead to a better understanding of protein folding.

1.2.3 Dynamics and Conformations in Protein-Chaperone Complexes

Due to possible transient population of non-native structures of proteins during synthesis and folding within a living cell, there is a significant risk that these structures expose hydrophobic surfaces and therefore tend to aggregate. There is increasing evidence that accumulation of protein aggregates may lead to pathologic phenotypes as the one found in neurodegenerative disorders such as Parkinson‘s disease and Huntington‘s disease [97]. During the last 20 years a remarkable number of proteins have been found which assist other proteins in finding their native conformation [117, 34]. According to a definition by Vabulas *et al.*, chaperones are defined as proteins which interact, stabilize or help non-native proteins in acquiring their final functional conformation, but are not present in the final functional structure [117]. The fact that chaperones are found in all three kingdoms of life underlines their evolutionary importance [34].

A remarkable representative of such a chaperone system is the bacterial chaperonin GroEL together with its co-chaperonin GroES. GroEL consists of two heptameric rings made from 57 kDa subunits. Two such rings are assembled back to back to the final barrel shaped GroEL. The co-chaperonin GroES is made of seven 10 kDa subunits which build up the lid structure that can be bound on top of GroEL. One GroEL ring together with the bound GroES encloses a cavity with a 65 Å diameter were substrate proteins can be encapsulated [114]. Since GroEL accounts for roughly 1% of the soluble cytoplasmic protein in bacteria, there is an ongoing debate about the number of possible substrate proteins [44]. Whereas experiments by Vitanen *et al.* [118] showed that up to 50% of the soluble proteins of Escherichia coli can be bound to GroEL after denaturation, there are more recent results by Houry *et al.* and Kerner *et al.* which show that there are 250-300 strong binding substrate proteins [45, 54]. The asymmetric reaction cycle of this chaperone is shown in figure 2.

While the catalytic cycle is understood rather well, the influence of this machinery on the folding pathway of its various substrates remains a controversial issue [42]. For more information on how the GroEL/ES chaperonin system may affect the folding energy landscape of its substrate proteins, see section 1.2.4.

To get new insights in the GroEL/substrate interaction, we started to study the initial bound state of a substrate protein which is illustrated in figure 2(b). Because it is difficult to study the GroEL chaperone system under turnover conditions due to the possibility of substrate protein binding to both rings, we decided to use a single-ring variant of GroEL (SR1), which binds to the substrate in a 1:1 complex. This variant carries mutations in the interface area between the two rings of the wild-type GroEL which prevent the formation of the double ring system [43]. This SR1 variant is perfectly capable of binding ATP and GroES but since there is no allosteric release signal from the opposite ring, the substrate will be trapped within the cavity [42]. Bovine liver rhodanese, a known model system for the GroEL/GroES system, was used as a substrate [68]. Rhodanese is a 293 amino acid mitochondrial protein, it consists of two globular domains of similar size connected by a 15 amino acid linker. Although there is a striking structural similarity between the two domains, no significant sequence homology has been found [86]. Refolding rhodanese in the absence of chaperones or tensides leads to significant aggregation even at low protein concentration

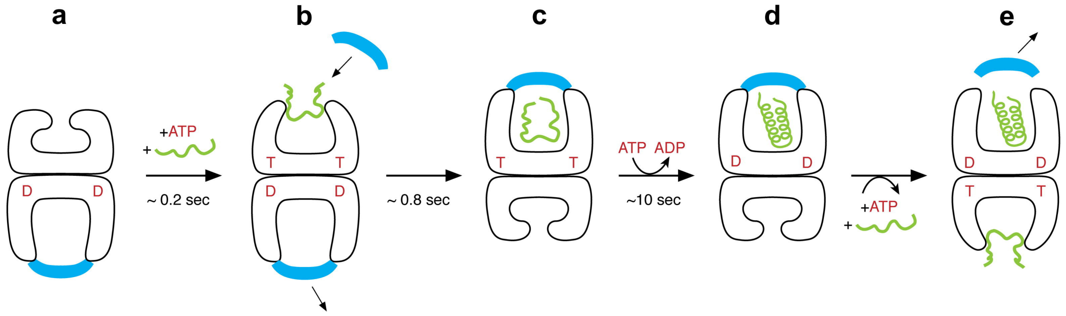


Figure 2: ATP-driven reaction cycle of the GroEL/ES chaperone system. The tetradecameric GroEL is represented in black. The heptameric co-chaperonin GroES is drawn in blue and the substrate protein in green. ATP binding is shown in red, where T stands for ATP and D for ADP. **(a)** Starting state with GroES and ADP bound to one subunit and the hydrophobic rim presented on the other. This hydrophobic rim captures non-native substrate proteins, and binding of ATP results in a small but cooperative conformational change which allows binding of GroES. **(b)** Binding of GroES leads to a large conformational change which switches the interior of the GroEL/ES cavity from hydrophobic to hydrophilic, which may contribute to substrate folding. This folding active state is called the *cis* complex. **(c)** Hydrolysis of ATP is the rate limiting step in the reaction cycle. Hydrolysis causes an allosteric change which triggers substrate binding in the opposite, so-called *trans* ring. **(d)** Binding of a new substrate protein and ATP to the trans ring and release of GroES and the former substrate protein from the *cis* complex. **(e)** The cycle is complete and the system is ready for the next turnover (Adapted from Horwich 2009 [42]).

[111, 36].

Complementary double cysteine variants, in which the primary vs. tertiary structure inter-dye distance was varied, were produced, and the cysteines were labeled with Alexa Fluor 488 as a FRET donor and Alexa Fluor 594 as an acceptor fluorescent dye. Double labeled rhodanese was unfolded in GdmCl and diluted into GdmCl-free buffer containing a stochiometric excess of unlabeled SR1, leading to quantitative binding of the rhodanese to SR1, as confirmed by size exclusion chromatography. FRET histograms of both variants showed an unusually broad distribution of transfer efficiencies for the SR1-bound state. The slight but significant difference in the shape and position of the FRET histograms for the two complementary variants of SR1-bound rhodanese implies that these broad histograms can report on the conformation of the SR1 bound rhodanese. Changing the unfolding strategy for rhodanese from GdmCl to a pH jump using phosphoric acid had no effect on the shape of the SR1-bound histograms. From this finding we concluded that the broad shapes do not reflect the distance distribution of the unfolded state but are rather the signature of an early folding intermediate. Extending the FRET fluorescence correlation spectroscopy (FCS) experiments using a Hanbury Brown and Twiss setup for freely diffusing molecules [74] to experiments with surface immobilized SR1-bound rhodanese allowed us to probe inter-dye dynamics of the bound rhodanese from ns to hundreds of ms. In addition to the expected photophysical effects such as photon-antibunching and triplet state dynamics, a component with a time constant of around 200 ns was observed. This 200 ns component could be attributed to rotational motion of the entire SR1-rhodanese complex (See section 1.1.1). Together with the polarization decays measured in this experiments, we concluded that there is no change in distance between the two reporting fluorescent dyes on all accessible timescales from ns to ms.

The remaining question was about the structural origin of the broad distribution in transfer efficiency histograms. Since no inter-dye dynamics were found, there were only two possible explanations for the broad histograms. Either, there was a static distribution of intermolecular distances or a static distribution of relative donor-acceptor fluorophore orientations. A comparison of fluorescence anisotropy decay measurements of free and unfolded rhodanese versus SR1-bound rhodanese revealed that the anisotropy in the SR1 bound state is significantly increased. The measured anisotropy decay times revealed a significant slow rotational component which means that the ps-ns rotation of the fluorescent dyes on their linkers is restricted upon binding to SR1. A Monte Carlo simulation, based on a model in which the relative orientation of the dyes within a single molecule is constrained to a wobbling cone but randomly distributed from molecule to molecule, was performed. Such a simulation resulted in anisotropy decays similar to those determined in the ensemble experiments. We extended the simulation to include normally distributed inter-dye distances and adjusted the mean and standard deviation to get the best match with respect to the experimental transfer efficiency histograms. In doing this for the two differently labeled rhodanese variants we were able to derive a possible binding conformation of rhodanese on the SR1 ring. One can conclude that at least one domain of rhodanese seems to be in a compact, probably native-like, conformation, whereas the binding to the SR1 happens at the interface area between the two rhodanese domains.

In these experiments we were able to get insights into protein conformations with single-molecule FRET in the presence of restricted chromophore flexibility. Additional improvements allowed us to extend the accessible timescale for inter-dye dynamics measurements into hundreds of ms with ns time resolution. This represents a 100 fold increase with respect to the diffusion time of such molecules trough the confocal volume, which is usually the natural upper time limit of such fluorescence correlation spectroscopy (FCS) experiments. In all the ns-FCS experiments performed on GroEL-bound and GroEL/ES-encapsulated rhodanese (See also section 1.2.4), there was always a significant rotational component observed but no additional contribution of chain dynamics within the experimentally accessible time-range of nanoseconds to seconds. Therefore, the conclusions from these experiments are that there is either no measurable rhodanese chain dynamics upon binding to the chaperone, or the steric restriction of the fluorescent dyes prevents the reporting of such a dynamic component, and the signal is completely dominated by the rotational effects. Together with ensemble lifetimes measurements and simulations, it was possible to gain new insights into the conformation and the dynamics of SR1-bound rhodanese.

1.2.4 Protein Folding Kinetics within a Chaperonin Cage

How the GroEL/ES system does assist substrate proteins in folding is currently one of the most important questions in the chaperonin field [44]. The proposed concepts are active unfolding, prevention of aggregation and active folding. The active unfolding hypothesis is based on an iterative annealing mechanism, in which monomeric proteins fold by a kinetic partitioning mechanism where the free energy landscape of folding consists of several minima, separated by energy barriers. Depending on the barrier heights, a folding protein is capable of exploring the energy landscape without assistance. However, proteins with high folding barriers may be trapped for longer timescales in non-native low-energy minima. Since such non-native conformations often expose hydrophobic surfaces, these proteins may represent GroEL substrates. After being bound to GroEL, such substrates experience a significant allosteric change of the GroEL upon binding of ATP and GroES, in which the GroEL cavity surface changes from hydrophobic to hydrophilic and the cavity volume expands twofold. It is believed that such an rearrangement overcomes the free energy barrier of the substrate and allows it to continue the search for its native structure. By following such a binding, stretching and release reaction scheme multiple times, the substrate is supposed to find its final energy minima and therefore the native state [114]. Another proposed mode of GroEL/ES action is the prevention of aggregation, which is summarized in the passive “Anfinsen folding cage” effect. No direct advantage on the folding pathway is expected from this mechanism, but since the folding substrate is confined within the cavity, this allows the substrate to find its native state in isolation and thus prevents aggregation [42]. Finally, there are studies suggesting that the volume and the charge distribution of the inner surface of the GroEL/ES cavity and its change from hydrophobic to hydrophilic does have a direct positive effect on the folding energy landscape of the substrate protein [112]. Additionally, simulations on different substrate peptides within confined volumes of different sizes and shapes show that the confinement found in the GroEL/ES chaperonin may lead to an acceleration of substrate protein folding due to a decrease in the entropy of the unfolded state compared to the transition state [56, 109]. However, doing explicit solvent simulations on villin headpiece within a confined volume showed that there is an opposing effect by confining the folding substrate together with water within such a cavity. Due to the hydrophobic containment within the cavity, water molecules can participate in less hydrogen bonds. Optimizing the hydrogen bond network between water molecules leads to a decrease of the folded state stability [66]. The controversy over how such an ubiquitous system acts and the experimental details accessible using single-molecule and classical biophysical methods made this an intriguing system for in-depth study. Therefore we decided to rigorously study the folding kinetics of a substrate protein within the GroEL/ES cavity.

To avoid the release of the folded substrate from the GroEL/ES cavity, we used the single-ring variant SR1. As a substrate, bovine liver rhodanese was used (See section 1.2.3). To probe for folding cooperativity and for optimal discrimination of native and nonnative conformations, three different FRET variants of rhodanese were produced as shown in figure 3.

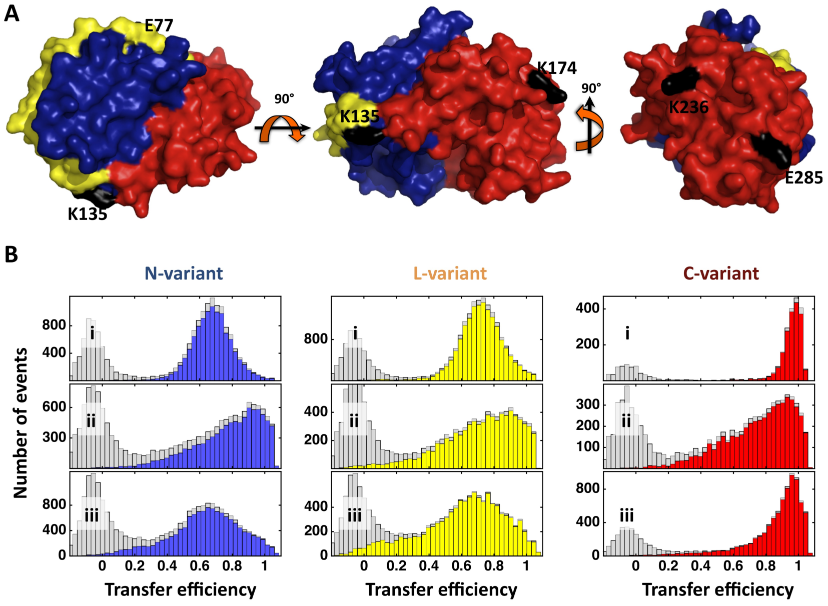


Figure 3: Native structure and transfer efficiency histograms of the rhodanese variants. **(A)** Surface representation of rhodanese showing the N-terminal domain (blue), the interdomain linker (yellow), and the C-terminal domain (red) (protein data bank entry 1rhs). The rhodanese variants were labeled with Alexa Fluor 488 as a donor and Alexa Fluor 594 as an acceptor. Label attachment sites are indicated in black. **(B)** Transfer efficiency histograms of native rhodanese (i), the SR1-rhodanese complex (ii), and the SR1-rhodanese complex 1.5 h after addition of GroES and ATP (iii). The gray histograms were recorded with donor excitation only; the colored histograms were recorded using dual color excitation of donor and acceptor to eliminate the contribution close to E = 0 from molecules lacking an active acceptor dye (Taken from [41]).

In the N-terminal and C-terminal variant, both fluorescent dyes were placed in the respective domain and therefore reported on the folding of each domain. In the linker or L-variant there was a dye in each domain reporting on the folding of the overall protein or more specificly on the arrangement of the interface region between the two domains. Binding to GroEL resulted in characteristically broad transfer efficiency histograms for all three variants (See section 1.2.3 and figure 3). Adding ATP and GroES to such GroEL bound rhodanese resulted in stable complexes with the rhodanese encapsulated within the chaperonin cavity. Over time the transfer efficiency histograms approached the shape of the corresponding native rhodanese with a residual broadening due to the restricted rotational freedom of the fluorescent dyes within the cavity. This change of the transfer efficiency histograms allowed us to monitor the folding kinetics within the cavity and therefore to compare the resulting rates to those of rhodanese folding free in solution. A combination of microfluidic and manual mixing single-molecule experiments allowed us to observe these folding kinetics from milliseconds to hours. Since subtle changes in such broad transfer efficiency histograms are difficult to fit due to the lack of an appropriate model, a model free singular value decomposition (SVD) method was used. To further improve the fitted kinetics, most of the available information from the single-molecule experiments such as burst-duration, photon rate, fluorescence lifetimes and anisotropies have been used in the SVD fit.

For the folding of rhodanese free in solution, similar rate constants for the N-terminal domain variant and the L-variant were found. In contrast, the C-terminal variant folded almost six times faster than either L or N variant. The same folding hierarchy was found for rhodanese encapsulated within the GroEL/ES complex. Whereas the rates for the N-terminal and the L-variant were not changed upon encapsulation, there was a twofold decrease in the folding speed of the C-terminal variant. To further elucidate the initial process of the chaperonin mechanism, microfluidic experiments in which GroEL bound rhodanese was mixed with GroES and ATP were carried out. These ms resolution single-molecule experiments revealed almost no change in transfer efficiency upon encapsulation. Interestingly, ensemble stopped flow experiments by Lin *et al.* on rubisco as a GroEL/ES substrate showed a change in transfer efficiency which was attributed to forced unfolding [64]. Calculating the average transfer efficiency from our single-molecule experiments and fitting the resulting values with two exponentials resulted in rate constants very similar to those reported in the literature for GroES binding and apical domain movement under substrate load. To get further insights into the deceleration mechanism of the C-terminal domain within the chaperonin, temperature dependent measurements of folding kinetics of all three variants free in solution and encapsulated in GroEL/ES were performed. The results of these experiments indicated that there is a significantly lower activation enthalpy of folding for the encapsulated C-terminal variant compared to its folding free in solution. Since such a behavior should increase the folding speed of the substrate, the enthalpy change can not explain the slower folding speed of the encapsulated rhodanese. The two other components which contribute to the height of the free energy barrier are the pre-exponential factor and entropy. Since conformational entropy of the substrate is reduced in a confined volume such as the GroEL/ES cavity, the entropy change between the unfolded and the folded state should be reduced, which will reduce the height of the free energy barrier. This should increase the folding speed within the GroEL/ES cavity. To test the contribution of solvent entropy, folding experiments in D_2O were performed. An almost two-fold decrease in folding speed was found for the C-terminal and the N-terminal domains of rhodanese free in solution and encapsulated within the GroEL/ES. The lack of a difference between autonomous and SR1-mediated folding upon changing the folding environment from H_2O to D_2O makes confined water molecules an unlikely cause for the change in folding rates upon encapsulation. The remaining component is the pre-exponential factor. It accounts for the roughness of the underlying free energy surface for folding. Possible contributing effects are nonnative interactions within the folding polypeptide chain or between the chain and the walls of the GroEL/ES cavity. Compared to the R_g of free, unfolded rhodanese, the available space within the GroEL/ES cavity is significantly smaller. Therefore, non-native interactions within the folding rhodanese chain are expected to be increased which in turn should increase the roughness of the energy landscape. However, since only the kinetics of the C-terminal domain is affected by the encapsulation of the whole rhodanese chain, there is reasonable doubt about such a conclusion. The only remaining contribution are interactions between the unfolded rhodanese chain and the cavity surface of GroEL/ES. Microfluidic mixing experiments show almost no change of the transfer efficiency histograms of SR1-bound rhodanese upon binding of ATP and GroES which indicates that there are similar interactions with the cavity walls of SR1 as in the bound state. A significant contribution from the change in surface conditions within the GroEL/ES upon encapsulation is rather improbable from these results. Dissociation from the wall of the cavity followed by a fast folding reaction might therefore be the rate-limiting step of the folding of the C-terminal domain of rhodanese within the GroEL/ES cavity.

Given that two domains within the same protein are affected differently, these experiments imply that there is no universal mechanism on how the GroEL/ES chaperone system affects the folding of a substrate protein. Structural and dynamic information with nm-resolution and accessible timescales from milliseconds to hours without common problems such as protein aggregation and signal averaging revealed that there is no speedup of protein folding within the GroEL/ES cavity. From the results found in this work, in which a slowdown of folding and no active unfolding is found for the GroEL/ES encapsulated rhodanese, the only remaining model which partly describes the function in our case is the passive “Anfinsen folding cage”. However, the observed effects which even vary between different domains of the same protein highlight the need of a general mechanistic reconsideration of the GroEL/ES chaperonin function.

1.3 Extensions and Applications of Single-Molecule Methods to Study Protein Chain Dynamics and Conformations in Complex Protein Systems

1.3.1 Development of a Rigorous Photon Statistics Model for a Description of Fluorescent Dye Photophysics and Protein Chain Dynamics

Chapter 6 describes the development and successful application of a novel photon statistics fitting approach to describe ns-FCS data. The pioneering approach by Gopich *et al.* [27, 28] and Nettels *et al.* [74] to describe ns-FCS experiments using detailed kinetic descriptions of the involved fluorescent processes was extended. All of the contributing effects, such as donor-only populations, due to incomplete labeling and photo bleaching, or non-perfect detection systems were included in the analysis. Additionally, the photon statistics fits were extended to include dynamic and static quenching of the involved fluorescent dyes. Using such a description allowed to fit all the available ns-FCS globally. This approach was successfully applied and benchmarked using various experimental systems. First, stiff polyprolines with and without a quenching Trp on a flexible linker were used to validate the fitting approach. The successful description of the resulting ns-FCS allowed us to increase the complexity and the inclusion of FRET dynamics between the two fluorescent dyes. As a test system for this scenario, novel variants of the N-terminal domain of HIV-1 integrase, an intrinsically disordered protein, were developed and measured extensively using ns-FCS. Extending the fitting approach to include diffusive inter-dye dynamics allowed us to quantitatively fit the resulting data. Together with various controls, this showed the general applicability of this approach to analyze ns-FCS data. It was possible to quantify and separate the various effects included in the ns-FCS data such as the FRET dynamics between the donor and the acceptor fluorescent dyes together with the PET quenching between the fluorescent dyes and a quenching Trp. The future application of this approach will allow us to address challenging questions about protein folding and dynamics in complex molecular systems.

1.3.2 Experimental Model Peptides for FRET and Quenching Dynamics

For the development and verification of single-molecule FRET methods, there is a continuing need for molecular test systems with clearly defined distances between the donor and the acceptor fluorescent dyes. Possible candidates are poly-L-prolines, such as those used by Stryer and Haugland in their pioneering experiments on distance dependent FRET [106] or double stranded DNA constructs such as those introduced by Cardullo *et al.* [9]. Both approaches were successfully applied in ensemble and single-molecule FRET experiments [106, 9, 30, 95]. Starting with the single-molecule FRET experiments of Schuler *et al.* on polyprolines of various lengths labeled with Alexa 488 and Alexa 594 fluorescent dyes [95], there has been increasing interest in the FRET distances and distributions obtained by such proline systems in single-molecule experiments. More refined experiments and simulations were performed to explain the origin of unexpected FRET heterogeneities and transfer efficiencies [96, 5].

In the work presented here, polyproline peptides were used for 3 different projects. As a reference sytstem for FRET measurements within the GroEL/ES cavity, a polyproline with 11 prolines and with a GroEL affinity tag to allow binding and encapsulation of this spectroscopic ruler within the chaperonin cavity was produced and successfully measured (Data not shown). Next, improved purification and measuring strategies were applied to obtain high resolution single-molecule data on FRET polyprolines of various lengths. This data was used to develop and verify a novel approach to simulate and fit experimental FRET efficiencies and distributions. This approach is based on the following procedure: First, multiple, fully atomistic simulations on the solvated polyprolines including the parametrized fluorescent dyes were performed. From the relative dipole orientations of the dyes in the simulations, time-dependent energy transfer rate coefficients are derived. Using this information, in the third step, a Monte Carlo simulation is performed to simulate and collect individual photon absorption, excitation, FRET transfer and emission events for randomly chosen points of the MD trajectory. Using an experimentally determined burst size distribution, these “emitted” photons are collected into bursts and analyzed as described for a standard single-molecule experiment (See section 1.1). The simulations revealed that even on rather simple systems such as polyprolines there are significant deviations due to a non-isotropic distribution of dye transition dipoles. This effect leads to a breakdown of the $\kappa^2 = 2/3$ assumption which is often assumed for the analysis of FRET data. Additionally, a slow transition between a freely rotating dye conformation and one where the dye sticks to the polyproline, mediated by hydrophobic interactions, was found in the simulations. An extensive experimental search for such an effect using fluorescence anisotropy decay experiments with various solvent conditions to change the extent of hydrophobic interactions was not able to confirm such a finding. From this, one can conclude that the contribution of this effect to the experimentally accesible observables is rather small or may be an artefact of the simulations. However, these experiments show the need to extend and optimize molecular dynamics force fields to simulate peptides together with the attached fluorescent dyes. Taken together, this work shows how a combined approach using single-molecule FRET experiments and simulations can give new insights into the mechanistic origins of the transfer efficiency distributions. It higlights the importance of taking dye orientation and flexibility, which are often constrained upon linking to a protein system, into account [38]. Additionally it also indicates the necessity for improved force field parametrization of fluorophores for future simulations.

In chapter 6, polyprolines where used to study the effect of dynamic and static quenching on transfer efficiencies and ns-FCS measurements. For this study, polyproline constructs with four C-terminally attached $(\text{Gly-Ser})_\text{n}$ repeats and a quenching tryptophan residue where produced and labeled with Alexa 488 an Alexa 594. Site specific labeling together with

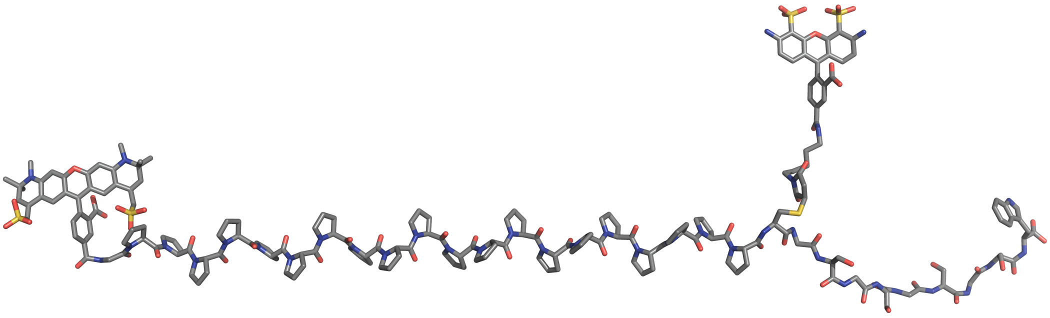


Figure 4: Structural representation of the polyproline construct used to measure the effect of PET quenching on single-molecule FRET experiments. Here, a 20 residue polyproline with a C-terminal (Gly-Ser)$_4$-Trp is shown. The cysteine close to the Trp-quencher is labeled with Alexa 488 as a FRET donor. The N-terminal glycine is labeled with the Alexa 594 acceptor dye.

selective purification of the labeling isomers allowed us to obtain peptides with the cysteine position close to the tryptophan labeled either with a donor or an acceptor dye. Varying the linker length between the fluorescent dye and the maleimide which reacts with the cysteine allowed us to study the effect of the fluorophore linker on transfer efficiencies and quenching dynamics. Single-molecule and ensemble FRET experiments were performed to measure the contribution of static and dynamic quenching on either the donor or the acceptor dye with respect to the distance of the tryptophan quencher. Based on these results, photon statistics models were developed and applied to fit auto- and crosscorrelation ns-FCS curves globally. These novel strategies to fit ns-FCS data allowed us to deconvolute PET-quenching and FRET processes.

1.3.3 Using PET and FRET in an Unified Approach to Measure Long and Short-Range Chain Dynamics in a Single Molecule

For the final part of this work (See section 6.4), novel variants of the N-terminal domain of HIV-1 integrase were produced. The constructs of this intrinsically disordered protein were designed such that they have two labeling positions for fluorescent dyes. One of the positions was chosen to be close to the intrinsic Trp and therefore allowing quenching of the fluorescent dye. The second position was on the C-terminus of the protein, as far away from the quenching Trp as possible. A rigorous labeling and purification process produced site-specifically labeled proteins. The intrinsically disordered nature of these protein constructs allowed us to measure unfolded chain dynamics using FRET between the two fluorescent dyes and PET using one of the dyes and the Trp residue under native conditions and in the same experiment. The application of the novel photon statistics fitting approach allowed us to quantitatively describe the long-range inter-dye dynamics together with the short-range PET quenching between the donor fluorescent dye and the quenching Trp. This showed for the first time the feasibility of separating and quantifying such dynamic effects in ns-FCS measurements. Various controls such as Trp free variants and donor-only constructs were produced and analyzed to ensure the validity of these results. A comparison of the FRET and the PET quenching dynamics revealed different behavior upon increasing GdmCl concentration. Whereas the long-range inter-dye reconfiguration time decreases, the local quenching association time between the dye and the Trp increases. This could indicate that there is an opposite behaviour between local and global chain dynamics in an unfolded protein upon denaturation with GdmCl. This preliminary finding indicates how this unified approach consisting of state of the art protein design and labeling together with photon statistics fitting may lead to novel insights about protein folding and dynamics.

2 Single-Molecule Spectroscopy of the Temperature-Induced Collapse of Unfolded Proteins

D. Nettels, S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, and B. Schuler.

Contributions of Dominik Hänni:

I developed an automated setup for performing single-molecule experiments using wellplates, which allowed performing large-scale screens of chemical substances and substance mixtures for photo-protection and photo-enhancement of the used fluorescent dyes. New data analysis strategies allowed the identification of the most efficient substance combinations in a set of thousands of measurements. The applications of such chemicals lead to an increased stability of the fluorescent dyes, especially at increased temperatures and/or increased laser powers. This allowed the successful measurement of the temperature-induced collapse of unfolded proteins. Not using such a chemical photo-protection system leads to a loss of the single-molecule fluorescence signal at increasing temperatures.

Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins

Daniel Nettels[a], Sonja Müller-Späth[a], Frank Küster[a], Hagen Hofmann[a], Dominik Haenni[a], Stefan Rüegger[a], Luc Reymond[a], Armin Hoffmann[a], Jan Kubelka[b], Benjamin Heinz[c], Klaus Gast[c], Robert B. Best[d], and Benjamin Schuler[a,1]

[a]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; [b]Department of Chemistry, University of Wyoming, Laramie, WY 82071; [c]Physikalische Biochemie, Universität Potsdam, 14476 Potsdam, Germany; and [d]Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, United Kingdom

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved September 25, 2009 (received for review March 6, 2009)

We used single-molecule FRET in combination with other biophysical methods and molecular simulations to investigate the effect of temperature on the dimensions of unfolded proteins. With single-molecule FRET, this question can be addressed even under near-native conditions, where most molecules are folded, allowing us to probe a wide range of denaturant concentrations and temperatures. We find a compaction of the unfolded state of a small cold shock protein with increasing temperature in both the presence and the absence of denaturant, with good agreement between the results from single-molecule FRET and dynamic light scattering. Although dissociation of denaturant from the polypeptide chain with increasing temperature accounts for part of the compaction, the results indicate an important role for additional temperature-dependent interactions within the unfolded chain. The observation of a collapse of a similar extent in the extremely hydrophilic, intrinsically disordered protein prothymosin α suggests that the hydrophobic effect is not the sole source of the underlying interactions. Circular dichroism spectroscopy and replica exchange molecular dynamics simulations in explicit water show changes in secondary structure content with increasing temperature and suggest a contribution of intramolecular hydrogen bonding to unfolded state collapse.

FRET | polymer | protein folding | secondary structure | chain dimensions

There is an increasing interest in the properties of unfolded proteins and their roles in the folding and cellular functions of proteins. A key motivation is that many proteins are marginally stable and only fold in the presence of their ligands or binding partners, opening new regulatory possibilities (1, 2). An important reason for recent progress is the growing availability of methods that provide structural information on these conformationally heterogeneous systems, such as NMR (3), scattering methods (4, 5), and single-molecule FRET (6–8). Although NMR provides mostly local details, small-angle X-ray scattering (SAXS), dynamic light scattering (DLS), and single-molecule FRET provide overall hydrodynamic or long-range distance information. An important advantage of single-molecule FRET is the separation of folded and unfolded subpopulations (9). As a result, unfolded state properties can be investigated even in the presence of folded molecules (i.e., under near-native conditions, which are physiologically most relevant). This advance has led to the observation of a continuous collapse of the unfolded state with decreasing denaturant concentrations (10), a behavior that now has been demonstrated for a large number of proteins (11–18) and peptides (19). Recent advances in the application of theoretical models have led to a quantitative description of this unfolded state collapse in terms of polymer-physical concepts (15, 20–23). Such chain compaction also has been demonstrated to result in increased internal friction and a slowdown of intramolecular dynamics of the polypeptide (19, 26), which can affect the kinetics of protein folding (27). However, it is still unclear which interactions drive unfolded state collapse.

An experimental variable that can provide more information about the nature of these interactions is temperature. The temperature dependence of protein collapse also has a substantial bearing on several fundamental aspects in protein folding, especially on the degree of energetic frustration and, correspondingly, on folding kinetics (28–33). For simple polymers and polymer models that do not involve a temperature-dependent interaction energy, conformational entropy favors open conformations, leading to chain expansion with increasing temperature (34), an effect that is assumed frequently to be dominant also for unfolded proteins. Although none of the sparse experimental results on this topic show such an expansion, they exhibit substantial variation, ranging from the absence of any detectable temperature dependence (35, 36) to slight (37, 38) and stronger (39) collapse, with some inconsistencies between measurements even on the same protein (35, 37, 40), calling for a systematic investigation of this issue. Ensemble investigations are limited largely to highly unfolding conditions to exclude interference from the signal of folded molecules and to minimize aggregation. Although these limitations can be avoided in single-molecule FRET experiments (6–8), the analysis of changes in the conformational distribution of unfolded proteins has remained largely limited to the influence of denaturants. Possible reasons may be insufficient precision of previous single-molecule experiments or difficulties in accurate temperature control of the confocal optics used over a sufficiently broad temperature range. Here, we overcome these complications (Fig. S1) and use the combination of single-molecule FRET with several other biophysical methods to investigate the effect of temperature on unfolded state structure and dimensions over a wide range of denaturant concentrations.

Results

Single-Molecule FRET Experiments. We first used a variant of the small cold shock protein Csp*Tm* (10, 16, 26, 41) labeled with donor (Alexa Fluor 488) and acceptor (Alexa Fluor 594) dyes close to its termini (10, 16). The efficiency of energy transfer between the dyes upon donor excitation was determined from photon bursts originating from individual molecules freely diffusing through the focal spot of the laser beam as $E = n_A/(n_A + n_D)$, where n_A and n_D are, respectively, the number of acceptor and donor photons emitted by the molecule (including corrections; see *SI Materials and Methods*). A histogram from a large number of such events shows distinct maxima corresponding to the subpopulations present in the sample (Fig. 1). The peak at high E corresponds to folded molecules, and the peak at

---

Author contributions: D.N., S.M.-S., R.B.B., and B.S. designed research; D.N., S.M.-S., F.K., H.H., D.H., S.R., L.R., A.H., J.K., B.H., K.G., R.B.B., and B.S. performed research; D.N., S.M.-S., F.K., D.H., S.R., L.R., A.H., J.K., B.H., K.G., R.B.B., and B.S. contributed new reagents/analytic tools; D.N., S.M.-S., F.K., H.H., B.H., K.G., R.B.B., and B.S. analyzed data; and D.N., S.M.-S., K.G., R.B.B., and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: schuler@bioc.uzh.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900622106/DCSupplemental.

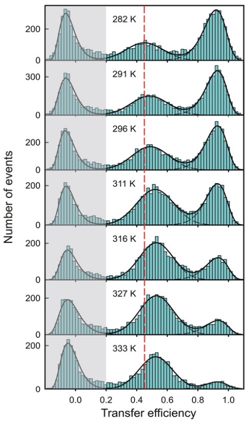


Fig. 1. Single-molecule FRET efficiency histograms of Csp*Tm* show thermal denaturation and a temperature-induced collapse of the unfolded protein (1.0 M guanidinium chloride). The peak at $E \approx 0.9$ corresponds to folded molecules; the peak at intermediate E corresponds to unfolded molecules. The peak at $E \approx 0$ (shaded) originates from molecules with an inactive acceptor. To determine mean transfer efficiencies, the unfolded peak was fit to a normal distribution, and the other two peaks were fit to log-normal distributions (black lines).

intermediate E corresponds to unfolded molecules. Fig. 1 shows an example of a temperature-dependent single-molecule FRET measurement at a guanidinium chloride (GdmCl) concentration of 1.0 M. With increasing temperature, we observe a decrease in the folded population and a concomitant increase in the unfolded population, as expected from the thermal denaturation of a two-state protein. However, we also observe an increase in the mean transfer efficiency of the unfolded subpopulation, $\langle E\rangle$, with temperature, corresponding to a collapse of the chain. Fig. 2*A* shows that this increase in $\langle E\rangle$ with temperature occurs at all GdmCl concentrations. The small change in transfer efficiency with temperature is probably the reason why we have not been able to detect a significant amplitude from unfolded state collapse in laser temperature-jump experiments, similar to experiments with a closely related cold shock protein (42). The transfer efficiency of unfolded Csp*Tm* at a constant temperature decreases with increasing GdmCl concentrations (Fig. 2*A Inset*), reflecting the well-established denaturant-induced expansion of the polypeptide (10–19).

For an analysis in terms of chain dimensions, the transfer efficiencies need to be converted to a measure of molecular size. Previous results indicate that on the length scales probed here, the distance distributions within unfolded Csp*Tm* can be approximated by those of a Gaussian chain (16). Taking into account the distance distribution $P(r)$, $\langle E\rangle$ in the unfolded state can be expressed as (10, 16)

$$\langle E\rangle = \int_a^{l_c} E(r)P(r)dr \Big/ \int_a^{l_c} P(r)dr, \text{ with } E(r) = 1/(1+(r/R_0)^6), \qquad \textbf{[1a,b]}$$

$$\text{and } P(r) = 4\pi r^2\left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2}\exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right), \qquad \textbf{[1c]}$$

where r is the distance between donor and acceptor and R_0 is the Förster radius. Given a value of $\langle E\rangle$ from the unfolded subpopulation, the mean-squared end-to-end distance $\langle r^2\rangle$ of the chain thus can be calculated numerically. To provide an intuitively accessible quantity, we plot our data in terms of the radius of gyration (Fig. 2*B*), which for a Gaussian chain is given by $R_g^2 = \langle r^2\rangle/6$ (34). Choosing a different polymer model for the analysis (e.g., that of a worm-like chain) does not affect our conclusions because of the similar shapes of the distance distributions in the range of persistence lengths relevant here (22). Note also that the effect of chain dynamics on the observed transfer efficiencies (43) does not affect our analysis significantly (Fig. S2).

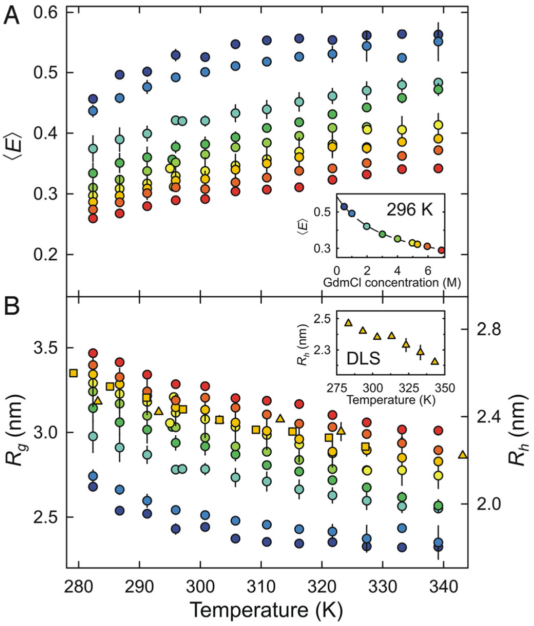


Fig. 2. Temperature dependence of (*A*) the mean transfer efficiency $\langle E\rangle$ and (*B*) the radius of gyration (R_g) of unfolded Csp*Tm* at different guanidinium chloride (GdmCl) concentrations (0.49, 1.0, 2.0, 3.0, 4.0, 5.0, 5.3, 6.0, and 6.9 M, from dark blue to red). Values from single-molecule measurements are shown as circles. The R_g values from ensemble fluorescence lifetime measurements in 5.3 M GdmCl are shown as orange squares, and the hydrodynamic radii (R_h) from dynamic light scattering (DLS) measurements in 5.3 M GdmCl are shown as orange triangles (right axis). The mean transfer efficiencies at 296 K (*inset* in *A*) illustrate the denaturant-dependent unfolded state collapse. For clarity, the inset in (*B*) shows the DLS results. Error bars indicate standard deviations estimated from two or three independent measurements for the cases where several measurements are available.

An ideal complementation of the distance information from FRET efficiency histograms, where the distance information is averaged over the duration of the fluorescence bursts, is the analysis of fluorescence intensity decays, which occur on a time scale of a few nanoseconds, much shorter than the reconfiguration time of the chain (26), and thus provide direct information about $P(r)$ (12, 16, 44). Additionally, fluorescence lifetimes are less susceptible to inaccuracies in instrument calibration. We thus performed ensemble time-correlated single photon counting experiments in 5.3 M GdmCl (Fig. 2*B*). The good agreement of the resulting values of R_g with the single-molecule measurements (Fig. 2*B*) confirms the accuracy of the correction factors (*SI Materials and Methods*) used for the determination of transfer efficiencies and distances from the E histograms. Additionally, we performed DLS experiments under identical conditions (5.3 M GdmCl) with unlabeled Csp*Tm*. Despite the less extensive statistics compared with the FRET experiments and some uncertainty from the slight protein concentration depen-

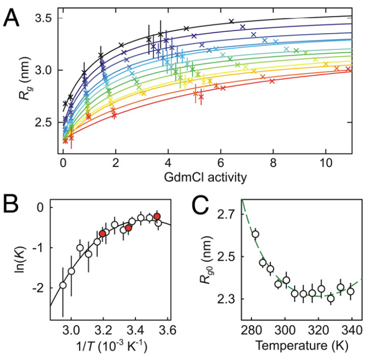


Fig. 3. Thermodynamic analysis. (*A*) The radius of gyration (R_g) of unfolded Csp*Tm* as a function of guanidinium chloride (GdmCl) activity at different temperatures (282, 287, 291, 296, 301, 306, 311, 316, 322, 327, 333, 339 K, from black to red crosses) with fits using Eq. **2** (lines). Error bars indicate standard deviations estimated from two to three independent measurements for the cases where several measurements are available. (*B*) The van't Hoff plot of the effective binding constants from the fits in *A* (open circles) compared with the values from calorimetric measurements taken from ref. 45 (red circles). (*C*) The radii of gyration at zero GdmCl activity (R_{g0}) from the fits in *A* show the temperature-induced collapse in the absence of denaturant. The dashed line is a fit to Eq. **3** ($T_0 = 295$ K, $\Delta H_0 = -1.3 \pm 0.1$ kJ mol^{-1}, $\Delta S_0 = -8.5 \pm 0.4$ J mol^{-1} K^{-1}, $\Delta C_p = 50 \pm 9$ J mol^{-1} K^{-1}). Error bars in *B* and *C* represent standard deviations taken from the covariance matrices of the fits in *A*.

dence of DLS measurements, the hydrodynamic radius R_h clearly decreases with temperature by ≈10 ± 3% between 10 and 70 °C, compared with a decrease of R_g by ≈15 ± 2% from FRET. The ensemble experiments under completely unfolding conditions thus agree with the single-molecule results, confirm the accurate calibration of the single-molecule instrument, and exclude a strong influence of the fluorophores on chain collapse. The agreement of single-molecule and DLS experiments suggests that the collapse of the unfolded state is a global process that affects the entire chain. Experiments with Csp*Tm* variants labeled at different positions (Fig. S3) support this conclusion.

Thermodynamic Analysis. In a first step toward the identification of the factors contributing to the temperature-induced chain collapse, we compare our results with the thorough calorimetric measurements of Makhatadze and Privalov on the interaction of denaturants with unfolded polypeptides (45). With increasing temperature, they found a decrease in the effective binding constants of urea and GdmCl to apo-cytochrome *c*, reduced and carboxymethylated ribonuclease, and lysozyme. The resulting temperature-induced dissociation of denaturant from the polypeptide is expected to lead to a collapse of the chain. Could this effect be sufficient to account for our observations?

A quantitative comparison requires the conversion of our results (Fig. 2) from GdmCl concentrations to thermodynamic activities *a* (45) to obtain the radius of gyration as a function of GdmCl activity at different temperatures (Fig. 3*A*).[‡] To allow a direct comparison with the calorimetric data, we use the same simple formalism as in ref. 45 to describe the effective binding of GdmCl to the unfolded polypeptide, assuming identical independent binding sites with an effective binding constant *K*:

$$R_g(a) = R_{g0}\left(1 + \rho\,\frac{Ka}{1 + Ka}\right), \qquad [2]$$

where R_{g0} is the radius of gyration at zero denaturant and ρ is the relative change in the radius of gyration approached asymptotically at very high GdmCl activities. A van't Hoff plot of the values of *K* obtained from our data upon fitting with Eq. **2** is in surprisingly good agreement with the calorimetric data (Fig. 3*B*).[§] With the extended temperature range available here, a curvature of the plot is apparent, resulting in a heat capacity change of $\Delta c_p = 0.5 \pm 0.1$ kJ mol^{-1} K^{-1} and an interaction enthalpy of $\Delta h = -9 \pm 3$ kJ mol^{-1} at 298 K compared with $\Delta h = -11 \pm 2$ kJ mol^{-1} from the calorimetric data (assuming $\Delta c_p = 0$) (45). This agreement corroborates the intimate connection between denaturant interactions and the collapse of the unfolded state that we observe (i.e., the temperature-induced dissociation of denaturant contributes to chain compaction). However, if this were the only contribution, we would expect the curves in Fig. 3*A* to converge at zero denaturant concentration, which is not the case: The values of R_{g0} obtained from the fits to Eq. **2** (Fig. 3*A*) clearly exhibit a temperature dependence (Fig. 3*C*) similar in shape to the temperature dependence at the lowest denaturant concentrations (Fig. 2*B*). We thus find a temperature-induced compaction of unfolded protein even in the absence of denaturant.

To clarify its molecular basis, we analyzed the data in Fig. 3*C* in terms of an extended form of Flory–Huggins theory (28), one of the simplest descriptions that can capture the effects of temperature on intramolecular interactions that lead to chain compaction

$$R_{g0} = \left(\frac{2nv}{3(2\chi - 1)}\right)^{1/3}, \qquad [3a]$$

where *n* is the number of monomers in the chain, *v* is the volume of a monomer, and χ is the Flory parameter, which is essentially a measure for the free energy of intrachain interactions relative to chain–solvent interactions (28). To describe the behavior of $R_{g0}(T)$, we introduce the temperature dependence of χ with

$$\chi(T) = \frac{1.4}{RT}\left(\Delta H_0 + \Delta C_p(T - T_0) - T\left(\Delta S_0 + \Delta C_p \ln\frac{T}{T_0}\right)\right) \qquad [3b]$$

where ΔH_0 and ΔS_0 are the enthalpic and entropic contributions, respectively, at T_0 and ΔC_p is the heat capacity contribution to χ (28). Interestingly, the fit to our data (Fig. 3*C*) predicts that the collapse will be followed by an increase in R_{g0} at high temperature. Unfortunately, this temperature regime is currently not accessible experimentally because of the rapid loss of signal >340 K, presumably due to accelerated photodestruction of the fluorophores.

Role of the Hydrophobic Effect. The interaction that most obviously may be assumed to be responsible for the temperature-induced collapse of unfolded proteins is the hydrophobic effect (39, 47). Even though we do observe a significant heat capacity term for the intrachain interactions from the analysis using Eq. **3** (Fig. 3*C*), the temperature dependence does not resemble the one typically assigned to the hydrophobic effect, with its more gradual increase and maximum strength approached only at ≈400 K (47). To clarify the role of the hydrophobic effect further, we thus performed single-molecule FRET experiments on prothymosin α, an intrin-

---

[‡] Note that for salts, such as GdmCl, the mean activity coefficient $\gamma_{\pm}$ of Gdm^+ and Cl^- is used. The dimensionless activity results from the molar GdmCl concentration *c* as $a = (\gamma_{\pm} c \text{M}^{-1})^2$. We used the tabulated values of the activity coefficient (46) for interpolation.

[§] The assumption made in Eq. **2** that the change in R_g is approximately proportional to the saturation of binding sites with denaturant molecules is plausible (45), even though we are not aware of a rigorous theoretical justification. However, the values of *K* obtained from fits analogous to Eq. **2** for either R_g (e.g., at 296 K, $K = 0.68 \pm 0.10$), the persistence length of the chain ($K = 0.58 \pm 0.07$), or R_g^3 (an effective molecular volume) ($K = 0.46 \pm 0.07$) as a function of GdmCl activity are all equal within error to the calorimetric value ($K = 0.60 \pm 0.09$ at 298 K), indicating that the choice of parameter is not particularly critical for our analysis.

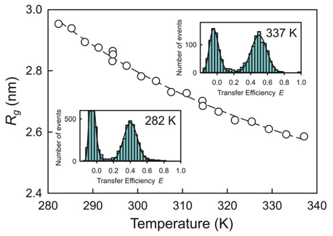


Fig. 4. Temperature-induced collapse of the intrinsically disordered protein prothymosin α. The apparent R_g calculated from $\langle r^2\rangle$ of the labeled protein segment is plotted. (*Insets*) Corresponding FRET efficiency histograms at 282 and 337 K are shown as examples. An empirical fit used for interpolation is shown as a dashed line.

sically disordered protein with extremely low hydrophobicity and an unusually large proportion of charged and polar amino acids (2, 48). In the case of a dominant contribution of the hydrophobic effect, we would expect a significantly less pronounced collapse for prothymosin α with increasing temperature than that for Csp*Tm*. We introduced Cys residues at positions 2 and 56 and labeled them with the same donor and acceptor fluorophores as Csp*Tm*. Single-molecule FRET experiments were performed in the presence of 0.5 M GdmCl to screen electrostatic interactions, which strongly influence the dimensions of prothymosin α at low ionic strengths due to its large negative net charge at pH 7 (48). Surprisingly, we observed the same degree of collapse for prothymosin α as for Csp*Tm*, with a reduction in R_g by ≈13 ± 3% between 282 and 337 K (Fig. 4) compared with 13 ± 2% for Csp*Tm* in 0.5 M GdmCl. A specific effect of GdmCl can be excluded, because a collapse of the same extent is observed for prothymosin α in the presence of other solutes that shield the charges by an increase in ionic strength (e.g., in 0.5 M sodium phosphate) (Fig. S4). These findings suggest that the hydrophobic effect is not the main cause of temperature-induced unfolded state collapse.

An aspect that has been implicated in the behavior of unfolded proteins is the formation of secondary structure with increasing temperature. Yang et al. suggested a general propensity of unfolded polypeptides to form local extended segments at high temperatures (49), visible in NMR (50) and as an increase in the CD signal around 222 nm (ref. 49 and citations therein). To investigate the role of this effect for Csp*Tm*, we used a destabilized protein variant to exclude the influence of the unfolding transition on the observed signal change. We prepared a deletion variant lacking the five C-terminal amino acids, resulting in complete unfolding even without denaturant. To eliminate problems with aggregation at high temperatures and for direct comparability with the collapse data, the experiments were performed in 0.5 M GdmCl. A comparison of the CD spectra at 277 and 368 K (Fig. 5) shows the characteristic signal change that indicates the formation of secondary structure (49). The reversibility of the process and the absence of aggregates after return to low temperatures show that aggregation is not involved, as also observed in previous studies (49, 51). Even though the temperature dependence of the increase in secondary structure content extends over a broader temperature range than the collapse observed by FRET (Fig. 3*C*), it is suggestive of a connection between chain collapse and secondary structure content.

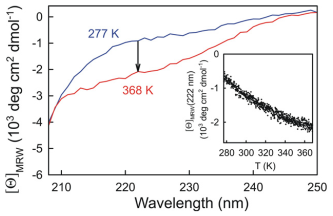


Fig. 5. Secondary structure content of unfolded Csp*Tm* increases with temperature. Circular dichroism spectra of Csp*Tm* destabilized by C-terminal truncation in 0.5 M guanidinium chloride at 277 K (blue) and 368 K (red). (*Inset*) Change in ellipticity at 222 nm is shown as a function of temperature.

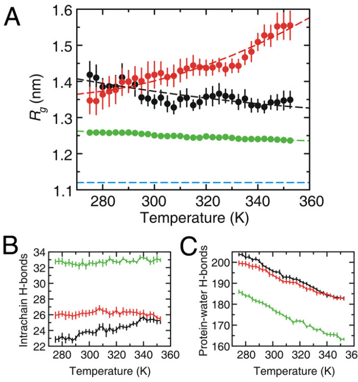


Fig. 6. Results on unfolded Csp*Tm* from replica exchange molecular dynamics simulations using AMBER ff03*/TIP4P-Ew water (black), AMBER ff03*/TIP3P water (red), and OPLS/AA-L/TIP3P water (green). (*A*) Radius of gyration (R_g) is shown as a function of temperature. Empirical fits used for interpolation are shown as dashed lines. The blue dashed line indicates R_g of the folded protein. Average numbers of intramolecular hydrogen bonds (*B*) within unfolded Csp*Tm* and (*C*) between protein and water molecules as a function of temperature are shown.

Simulations. To investigate further the microscopic origin of temperature-induced collapse, we performed replica exchange molecular dynamics simulations of unfolded Csp*Tm* in explicit water over a temperature range from 275 to 352.5 K (32 replicas) with different force fields and water models. Given that the experimental reconfiguration time of the unfolded state ranges between ≈20 and 80 ns under these conditions (26, 52) (Fig. S2), we can expect significant sampling of conformational space in the simulation time of 100–150 ns per replica (Fig. S5). Interestingly, we find that the results strongly depend on the force field and water model used. The AMBER ff03* protein force field (53) in combination with the TIP4P-Ew water model (54) resulted in a decrease of R_g by ≈6% over the entire temperature range (Fig. 6*A*), in qualitative agreement with the experimental result (Fig. 3*C*), but with TIP3P water (55), R_g increased by ≈14% (Fig. 6*A*). This difference may be related to the more accurate temperature dependence of the properties of pure water given by TIP4P-Ew compared with those given by TIP3P (54). Simulations with the OPLS-AA/L protein force field (56) also showed a slight collapse with temperature (Fig. 6*A*) but were less well converged (*SI Materials and Methods* and Fig. S6). In all cases, the mean R_g in the simulations (1.24–1.56 nm) was considerably lower than that observed experimentally [≈2.2–2.4 nm, corrected for the lengths of dyes and linkers (16)] These observations stress the importance of the subtle balance between protein and water energy functions (57), which is particularly critical for simulations of unfolded proteins.

Analysis of the backbone conformations (Fig. S7*a*) indicated an increase in the population of the α region of the Ramachandran map; the polyproline II content decreased with temperature, similar to NMR results for a short alanine peptide (50). A DSSP (58) analysis (Fig. S7*b*), which is largely based on hydrogen bonding patterns, showed an increased turn population at higher temperatures. Whereas the larger compactness in the simulations using OPLS compared with those using AMBER correlated with β-sheet content (Fig. 6*A* and Fig. S7*b*), the secondary structure changes observed with TIP3P and TIP4P-Ew were quite similar. The simulations thus may not allow us to identify a particular type of secondary structure involved in chain collapse (possibly also because longer time scales may be required to reach a completely equilibrated distribution of local secondary structure motifs), but we observed a clear overall correlation between collapse and intramolecular hydrogen bonding: in those force fields that collapsed the chain with increasing temperature, more protein–protein hydrogen bonds were formed and more protein–water hydrogen bonds lost than in the ff03*/TIP3P simulation, where the protein expanded with temperature (Fig. 6 *B* and *C*).

In summary, the simulations indicate roles for secondary structure formation and hydrogen bonding in temperature-induced collapse, but the strong dependence of the results on the water model indicate that the force fields—at least on the time scales accessible here—capture predominantly effects related to solvation.

Discussion

The few previously reported experiments give a heterogeneous picture of the effect of temperature on unfolded state dimensions. For example, although DLS measurements on RNase A unfolded by reduction of its disulfides showed a slight compaction with temperature (37), recent SAXS measurements reported no significant change of R_g (35). A significant collapse of RNase T1 in 5.3 M GdmCl was observed by DLS (38). No change in R_g was reported for β-lactoglobulin in 8 M urea between 0 and 25 °C (36). Laser temperature-jump experiments on the acid-denatured small protein BBL monitored by ensemble FRET showed a clear collapse with increasing temperature (39), and NMR indicated an expansion of the cold-denatured C-terminal domain of protein L9 with decreasing temperature (59).

Here, we tried to exclude the most important complications in the investigation of unfolded state collapse. First, we used a two-state folder without disulfides to avoid ambiguities from the influence of structured intermediates or disulfides on the properties of the denatured state. Second, we used single-molecule FRET to monitor the unfolded state even under conditions where the folded state is populated, avoiding the need for additional destabilization by reduction or low pH (35, 37, 39, 40). Third, aggregation does not occur at the exceedingly low concentrations used in single-molecule experiments and can be monitored in situ (60). Fourth, we excluded a significant influence of changes in chain dynamics on the observed FRET efficiency changes by measuring nanosecond intensity correlation functions (Fig. S2), which report on the time scales of chain reconfiguration (26). Finally, we directly compared the single-molecule results with ensemble DLS experiments on the same protein without labels and under identical solution conditions to exclude a strong influence of the FRET dyes on the collapse process.

Our results reveal two conceptually different contributions to the temperature-induced collapse of unfolded proteins. One of them is the enthalpic component of the interaction between protein and denaturant, resulting in the dissociation of denaturant from the polypeptide chain with increasing temperature (45). However, denaturant dissociation cannot account for the entire effect of temperature on unfolded state collapse, as demonstrated by the change in chain dimensions at very low GdmCl concentrations (Fig. 2) and extrapolated to zero denaturant (Fig. 3*C*).[‖] Interestingly, the hydrophobic effect, the interaction most commonly associated with protein stabilization (47), cannot be the sole source of this temperature-induced compaction, because we observe a collapse of a very similar extent for two proteins with very different average hydrophobicities, Csp*Tm* and the "natively unfolded" protein prothymosin α. Moreover, Csp*Tm* unfolds with increasing temperature, whereas the unfolded state compacts (Fig. 1), another indication that the interactions stabilizing the native state and those compacting the coil are not of the same physical origin.

Both CD measurements (Fig. 5) and molecular simulations (Fig. 6 and Fig. S7) (49) point toward a role for secondary structure formation and hydrogen bonding in temperature-induced unfolded state collapse. Even though a contribution from α-helical conformations (23) and β turns (61) should not be excluded, the contribution of β-structure formation may be particularly relevant. In kinetic synchrotron radiation CD experiments, we found earlier that collapsed unfolded Csp*Tm* exhibits a higher content of β structure than the protein at high denaturant concentrations (16). Magg et al. (42) showed that the collapse of a highly homologous cold shock protein is accompanied by the local extension of a short segment that forms a β strand in the folded structure. Most importantly, indications of an increase in β structure with increasing temperature have been reported for several proteins (ref. 49 and citations therein), similar to the behavior that we observe for Csp*Tm* (Fig. 5). From theoretical considerations and simulations (49, 62), there has been suggested to be a general entropic bias toward extended structure, because steric interference is minimized and side chains are allowed to adopt the largest number of configurations. This effect will be amplified at high temperatures. The resulting short and probably highly dynamic β segments may form nonnative hydrogen bonds within the unfolded polypeptide and contribute to chain compaction (19, 57). The same interactions may be responsible for the aggregation of unfolded proteins (49), a process that is particularly prevalent at high temperatures.

Relating the global compaction of unfolded proteins to local structure formation with further experiments, simulations, and theory will be interesting. Nuclear magnetic resonance, for instance, may allow an experimental determination of local structural propensities related to compaction (3, 59). By introducing systematic perturbations to explicit-solvent force fields, one may be able to identify the energetic contributions to unfolded state collapse. An alternative theoretical approach is the use of analytical models, such as structure-based free energy functionals (24) or coarse-grained simulations employing funneled energy landscapes (25). For example, in a recent study the relative contributions of native-like interactions and hydrophobic and electrostatic effects in the denatured state could be quantified (25). Such models could capture the temperature dependence of structure formation in the unfolded state if the interaction potential varied with temperature. Our results also may help to explain why many proteins exhibit two-state kinetics, since strong collapse is disfavored under folding conditions (33).

In summary, our observations bring together several independent findings that in fact may be connected closely: the temperature-induced collapse of unfolded proteins, the thermodynamics of protein–denaturant interactions, and the changes in the secondary structure propensity of unfolded polypeptides at high temperatures. The observation that the temperature dependence of attractive intramolecular interactions overcompensates for conformational entropy and leads to chain collapse with increasing temperature is in contrast to simpler systems, such as homopolymers, where the chain expands at high temperatures (28, 34). Finally, the arising picture suggests that the term "hydrophobic collapse" may be an incomplete description of the compaction of unfolded polypeptides at low denaturant concentrations and high temperatures.

[‖] From an analysis of the entire dataset with Eqs. 2 and 3, we find that in >1.5 M GdmCl denaturant, dissociation accounts for ≈50% of the chain collapse between 282 and 339 K. At lower GdmCl concentrations, the denaturant-independent component dominates.

Materials and Methods

Expression, purification, and labeling of full-length Csp*Tm* were performed as described in refs. 10 and 16. The truncated variant of Csp*Tm* and prothymosin α were purified using a hexahistidine tag. Single-molecule fluorescence and correlation experiments were performed using a MicroTime 200 confocal microscope (PicoQuant) essentially as described in refs. 16 and 26. Changes in refractive index and spectral properties of the dyes with GdmCl concentration and temperature were measured independently and taken into account for the calculation of R_0. The temperature was adjusted with a Peltier-controlled sample holder and calibrated using the temperature-dependent fluorescence lifetime of rhodamine B (63) (*SI Materials and Methods* and Fig. S1). We used a combination of additives (β-mercaptoethanol and cysteamine; see *SI Materials and Methods*) to reduce photodamage at high temperatures. Ensemble fluorescence intensity decays of the donor and acceptor were fit globally to the transfer rate distribution expected for a Gaussian chain by reconvolution with the instrument response function (16). Dynamic light scattering was measured as described in ref. 64. Circular dichroism measurements were performed at a protein concentration of 0.23 mg/mL. The absence of aggregates was tested by static light scattering in a fluorometer or by the lack of any slope in absorption measurements >350 nm. Replica exchange molecular dynamics simulations of unfolded Csp*Tm* were run using the OPLS/AA-L (56) and AMBER ff03* (53) force fields with two explicit water models [TIP3P (55) and TIP4P-Ew (54)], spanning a temperature range from 275 to 352.5 K in increments of 2.5 K. From unfolded configurations generated at high temperatures, simulations were run for 100–150 ns per replica (for a total of ≈9.3 μs). For details, see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank B. Wunderlich for finite element calculations, W.A. Eaton for access to laser temperature-jump instrumentation and discussion, G. Hummer, G. Makhatadze, D. Makarov, S. Takahashi, and D. Thirumalai for valuable comments and discussions, and A. Vartapetian and A. Evstafieva for the gift of a plasmid encoding prothymosin α. This work was supported by the Swiss National Science Foundation, the National Center of Competence in Research for Structural Biology, a Starting Researcher Grant of the European Research Council (to B.S.), and a Royal Society University Research Fellowship (to R.B.B.).

1. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
2. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
3. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14.
4. Millett IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv Protein Chem* 62:241–262.
5. Gast K, Modler AJ (2005) Studying protein folding and aggregation by laser light scattering. *Protein Folding Handbook*, eds Buchner J, Kiefhaber T (Wiley-VCH, Weinheim, Germany), pp 673–709.
6. Haran G (2003) Single-molecule fluorescence spectroscopy of biomolecular folding. *J Phys Condens Matter* 15:R1291–R1317.
7. Michalet X, Weiss S, Jäger M (2006) Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* 106:1785–1813.
8. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
9. Deniz AA, et al. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA* 97:5179–5184.
10. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
11. Magg C, Schmid FX (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein. *J Mol Biol* 335:1309–1323.
12. Laurence TA, Kong X, Jäger M, Weiss S (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc Natl Acad Sci USA* 102:17348–17353.
13. Kuzmenkina EV, Heyes CD, Nienhaus GU (2006) Single-molecule FRET study of denaturant induced unfolding of RNase H. *J Mol Biol* 357:313–324.
14. Tezuka-Kawakami T, Gell C, Brockwell DJ, Radford SE, Smith DA (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J* 91:L42–L44.
15. Sherman E, Haran G (2006) Coil–globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
16. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
17. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci USA* 104:1528–1533.
18. Hofmann H, Golbik RP, Ott M, Hübner CG, Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* 376:597–605.
19. Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103:12394–12399.
20. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci USA* 105:13403–13408.
21. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11:83–93.
22. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
23. Wang Z, Plaxco KW, Makarov DE (2007) Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. *Biopolymers* 86:321–328.
24. Shoemaker BA, Wolynes PG (1999) Exploring structures in protein folding funnels with free energy functionals: The denatured ensemble. *J Mol Biol* 287:657–674.
25. Weinkam P, Pletneva EV, Gray HB, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proc Natl Acad Sci USA* 106:1796–1801.
26. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
27. Cellmer T, Henry ER, Kubelka J, Hofrichter J, Eaton WA (2007) Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. *J Am Chem Soc* 129:14564–14565.
28. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 20:447–490.
29. Socci ND, Onuchic JN (1994) Folding kinetics of proteinlike heteropolymers. *J Chem Phys* 101:1519–1528.
30. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195.
31. Plotkin SS, Wang J, Wolynes PG (1996) Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top* 53:6271–6296.
32. Thirumalai D, Klimov DK (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 9:197–207.
33. Chahine J, Nymeyer H, Leite VBP, Socci ND, Onuchic JN (2002) Specific and nonspecific collapse in protein folding funnels. *Phys Rev Lett* 88:168101.
34. Grosberg AY, Khokhlov AR (1994) *Statistical Physics of Macromolecules* (AIP Press, Woodbury, NY).
35. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR (2007) Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 367:609–615.
36. Katou H, Hoshino M, Kamikubo H, Batt CA, Goto Y (2001) Native-like β-hairpin retained in the cold-denatured state of bovine β-lactoglobulin. *J Mol Biol* 310:471–484.
37. Nöppert A, Gast K, Müller-Frohne M, Zirwer D, Damaschun G (1996) Reduced-denatured ribonuclease A is not in a compact state. *FEBS Lett* 380:179–182.
38. Gast K, et al. (1997) Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: A dynamic light scattering study. *FEBS Lett* 403:245–248.
39. Sadqi M, Lapidus LJ, Muñoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100:12117–12122.
40. Sosnick TR, Trewhella J (1992) Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* 31:8329–8335.
41. Perl D, et al. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* 5:229–235.
42. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX (2006) Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol* 360:1067–1080.
43. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Brownian motion of ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers* 17:11–31.
44. Haas E, Wilchek M, Katchalski-Katzir E, Steinberg IZ (1975) Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proc Natl Acad Sci USA* 72:1807–1811.
45. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226:491–505.
46. Makhatadze GI, Fernandez J, Freire E, Lilley TH, Privalov PL (1993) Thermodynamics of aqueous guanidinium hydrochloride solutions in the temperature range from 283.15 to 313.15-K. *J Chem Eng Data* 38:83–87.
47. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
48. Gast K, et al. (1995) Prothymosin α: A biologically active protein with random coil conformation. *Biochemistry* 34:13211–13218.
49. Yang WY, Larios E, Gruebele M (2003) On the extended β-conformation propensity of polypeptides at high temperature. *J Am Chem Soc* 125:16220–16227.
50. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR (2002) Polyproline II structure in a sequence of seven alanine residues. *Proc Natl Acad Sci USA* 99:9190–9195.
51. Gast K, Zirwer D, Damaschun G (2003) Are there temperature-dependent structural transitions in the "intrinsically unstructured" protein prothymosin α? *Eur Biophys J* 31:586–594.
52. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B* 112:6137–6146.
53. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides. *J Phys Chem B* 113:9004–9015.
54. Horn HW, et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665–9678.
55. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
56. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487.
57. Bolen DW, Rose GD (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* 77:339–362.
58. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
59. Li Y, Shan B, Raleigh DP (2007) The cold denatured state is compact but expands at low temperatures: Hydrodynamic properties of the cold denatured state of the C-terminal domain of L9. *J Mol Biol* 368:256–262.
60. Hillger F, Nettels D, Dorsch S, Schuler B (2007) Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc* 17:759–765.
61. Perczel A, Fasman GD (1992) Quantitative analysis of cyclic β-turn models. *Protein Sci* 1:378–395.
62. Fitzkee NC, et al. (2005) Are proteins made from a limited parts list? *Trends Biochem Sci* 30:73–80.
63. Benninger RK, et al. (2006) Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal Chem* 78:2272–2278.
64. Gast K, Nöppert A, Müller-Frohne M, Zirwer D, Damaschun G (1997) Stopped-flow dynamic light scattering as a method to monitor compaction during protein folding. *Eur Biophys J* 25:211–219.

Supporting Information

Nettels et al. 10.1073/pnas.0900622106

SI Materials and Methods

Preparation and Labeling of Proteins. Cysteine residues were introduced by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes as described in refs. 1 and 2. Expression, purification, and labeling of full-length Csp*Tm* were performed as described in refs. 1 and 3. The truncated variant of Csp*Tm* was expressed with a cleavable hexahistidine tag to allow for rapid purification and to minimize nonspecific degradation of this unfolded protein. To this end, the gene was cloned from the vector pET21a into pET47b(+), and the sequence coding for the five C-terminal amino acids was deleted by successive quick-change mutagenesis steps. The protein was expressed in LB medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted, and DNA was digested. The cleared supernatant was loaded on a HisTrap column (GE Healthcare Bio-Sciences AB) in 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, and 4 M guanidinium chloride (GdmCl). After the 280-nm UV absorption signal reached the baseline, the column was washed with two column volumes of 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, and 10 mM imidazole, and a gradient from 10 to 500 mM imidazole was used to elute the His-tagged protein. Human rhinovirus 3C protease (containing a His tag) was added to a final concentration of 0.3 mg/mL, and after 12 h at room temperature, the cleavage reaction was dialyzed against 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, and 10 mM imidazole and applied to a HisTrap column. The cleaved Csp*Tm* without a His tag was collected in the flow-through and concentrated. Labeling was performed as for the full-length protein.

Prothymosin α also was expressed with a cleavable hexahistidine tag for rapid purification. The coding sequence for human prothymosin α was cloned from the vector pHP12 (4) into pET47b(+). Cysteine residues were introduced in positions 2 and 56 by site-directed mutagenesis (residue numbering is from Met in the protein sequence, excluding the 19-residue N-terminal purification tag). The protein was expressed in Terrific Broth medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted, and DNA was digested. The cleared supernatant was loaded on a HisTrap column in 20 mM Tris, pH 7.0, 100 mM NaCl, and 2 mM β-mercaptoethanol. A gradient from 40 to 500 mM imidazole was used to elute the His-tagged protein. Fractions were identified via SDS/PAGE, pooled, extracted with butanol, and precipitated with ethanol (4). Pellets were dissolved in 50 mM sodium phosphate, pH 7.0, reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP), and purified on a Superdex 75 gel filtration column (GE Healthcare Bio-Sciences AB) in 100 mM sodium phosphate, pH 7.0, 2 mM β-mercaptoethanol, and 0.01% Tween. Fractions containing the full-length protein were combined, extracted with butanol, and precipitated with ethanol. The pellet was dissolved in 4 M GdmCl and 50 mM sodium phosphate, pH 7.0, and the protein concentration was determined by bicinchoninic acid assay (BCA Protein Assay Kit; Pierce). Fluorophore labeling was performed at a protein concentration of $\approx$0.1 mg/mL with a threefold excess of Alexa Fluor 488 and Alexa Fluor 594; gel filtration was used to remove the free dye. The correct molecular mass of the labeled protein was confirmed by MALDI mass spectrometry. The donor-only labeled and acceptor-only labeled prothymosin α resulting from such random labeling does not interfere with single-molecule FRET measurements.

Single-Molecule Fluorescence Spectroscopy. Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) equipped with a continuous-wave 488-nm diode laser (Sapphire 488–100 CDRH; Coherent) and an Olympus UplanApo 60×/1.20W objective. Sample fluorescence was separated into donor and acceptor components using a dichroic mirror (585DCXR; Chroma Technology) and two final filters (ET525/50M and HQ650/100; Chroma Technology). Each component was focused onto an avalanche photodiode (SPCM-AQR-15; PerkinElmer Optoelectronics), and the arrival time of every detected photon was recorded. Samples of labeled protein were diluted to a concentration of $\approx$20 pM in 50 mM sodium phosphate buffer at the appropriate GdmCl (Pierce) concentration, individually adjusted to pH 7. Tween 20 (0.001%; Pierce) was added to prevent surface adhesion of the protein (1). To reduce data acquisition time and thus minimize damage to the chromophores at high temperatures, the photoprotective additives β-mercaptoethanol (200 mM) and cysteamine (5 mM) were included. As a result, the measurements could be performed at a laser power of 400 μW at the sample, resulting in significantly more intense fluorescence bursts and a reduction of the acquisition time from approximately 1 h to 15 min per measurement (for 5,000–10,000 identified bursts). Successive photons detected in either channel separated by $<$100 μs were combined into one burst. Identified bursts were corrected for background, differences in quantum yields of the donor and acceptor, different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation, as described in ref. 5. A burst was retained as a significant event if the total number of counts exceeded 70.

The Förster radius R_0 was corrected for the changes in solution conditions, which were dominated by the change in refractive index n with GdmCl concentration (2). The overlap integral (6) for Alexa Fluor 488 emission and Alexa Fluor 594 absorption was found to be independent of GdmCl concentration and temperature within error of the measurement. The change in n with temperature was measured refractometrically; it was found to be $<$1% over the temperature range used here and very similar to the refractive index change of water with temperature (7) at all GdmCl concentrations. Thus, n is dominated by the change in GdmCl concentration. Assuming the change of n with absolute temperature T and GdmCl concentration c_D (8) to be mutually independent, $n(T, c_D)$ was approximated by the following interpolation function

$$
\begin{aligned}
n(T, c_D) = {} & 1.3338 + 0.0174716\ \mathrm{M}^{-1}\, c_D - 0.0001855\ \mathrm{M}^{-2}\, c_D^2 \\
& + 7.80416 \times 10^{-6}\ \mathrm{M}^{-3} c_D^3 + 0.000262 \\
& (323.5 - 48.5\ \mathrm{e}^{0.02062(275 - T/K)} - T/K).
\end{aligned}
$$

Fluorescence lifetime experiments on Csp*Tm* singly labeled with the donor or acceptor show a negligible change of the fluorescence lifetime (and thus the quantum yield) with temperature. Correlation experiments (9) show that under the conditions used here chain reconfiguration times are at least approximately an order of magnitude greater than the donor fluorescence lifetime (Fig. S2) but small relative to the mean burst duration of $\approx$1 ms, justifying the use of Eq. **1** (main text) (10), which neglects the influence of chain diffusion on the observed transfer efficiencies (11). By numerically solving the diffusion equation in the potential of mean force corresponding to $P(r)$ for a Gaussian chain modified by a sink term to account for FRET (11), with the intramolecular end-to-end diffusion coefficients obtained from correlation experiments (9) as a function of temperature (Fig. S2), we can estimate the effect of the change in intramolecular diffusion on the change in the radius of gyration (R_g) over the accessible temperature range to be $\approx 10\%$ of the observed amplitude (i.e., less than the experimental uncertainty). Time-resolved fluorescence polarization anisotropy decays indicate rapid reorientation of the dyes during the fluorescence lifetime of the donor (12); an orientational factor κ^2 (6) of 2/3 thus can be used to calculate R_0 (2). The transfer efficiency of the unfolded state was independent of the laser power in the range of excitation rates used here.

A custom-built temperature-controlled sample holder employing Peltier elements and a digital temperature controller (TC2812-LAB12; Cooltronic) with a PT100 temperature sensor was used (Fig. S1). Because of the danger of damage to the microscope objective at high temperatures, the objective cannot be heated to the highest temperatures required here. The contact of the sample cell to the objective via immersion water thus results in a temperature gradient in the sample, and the temperature needs to be determined directly at the confocal volume. It was calibrated via the temperature-dependent fluorescence lifetime of rhodamine B (13) determined in a custom-built temperature-controlled ensemble time-correlated single photon counting instrument (12), where the temperature can be measured with a thermocouple directly in the sample whose temperature is kept uniform by stirring. Rhodamine B lifetimes from time-correlated single photon counting measurements in the confocal instrument under identical conditions then can be converted to temperature with an accuracy of approximately ± 2 K. The sample was excited at 470 nm with a 20-MHz repetition rate using a pulsed diode laser (LDH 470; PicoQuant); emission was observed at magic-angle settings using a z 582/15 filter (Chroma Technology). For measurements on labeled Csp*Tm* in 5.3 M GdmCl, ET525/50M and HQ650/100 filters (Chroma Technology) were used for the donor and acceptor emission, respectively. The decays were fit globally to the transfer rate distribution expected for a Gaussian chain by reconvolution with the instrument response function (FWHM of 80 ps) (2). Note that at the extremely low excitation rate of these ensemble experiments, the donor-only peak is absent in our Csp*Tm* preparation, as in single-molecule experiments at very low power (14) or measurements under flow (14), and therefore does not influence the analysis.

Dynamic Light Scattering. Dynamic light scattering was measured at a scattering angle of 90° and a wavelength of 532 nm using a custom-built apparatus, equipped with a diode-pumped, continuous-wave laser (Millennia IIs; Spectra-Physics) and an avalanche photodiode. Details of the detection and data processing procedures have been described elsewhere (15). The translational diffusion coefficients D were obtained from the measured autocorrelation functions using the program CONTIN (16). The translational diffusion coefficients D were converted to hydrodynamic radii via the Stokes–Einstein equation $R_h = k_B T/(6\pi\eta_0 D)$, where k_B is Boltzmann's constant, T is absolute temperature, and η_0 is the solvent viscosity. Solvent viscosities and densities were measured using an Ubbelohde type viscometer (Viscoboy 2; Lauda) and a digital density meter (DMA 58; Anton Paar), respectively. Refractive indices of the solvent mixtures were measured with an Abbe type refractometer. The solvents and protein solutions were filtered through 100-nm pore-size Anotop filters (Whatman) directly into small flow cells with path lengths of 1.5 mm (Hellma).

Circular Dichroism Spectroscopy. Experiments were performed in a Jasco J-715 CD spectrometer (Jasco) equipped with a temperature-controlled water bath, using a cylindrical quartz cell with a path length of 1 mm and a protein concentration of 0.23 mg/mL in 50 mM sodium phosphate buffer with 0.5 M GdmCl adjusted to pH 7.0. The absence of aggregates was tested by static light scattering in a fluorometer or by the lack of any slope in absorption measurements at wavelengths >350 nm.

Molecular Dynamics Simulations. Simulations of unfolded Csp*Tm* were run in explicit solvent using periodic boundary conditions with the GROMACS 4.0.3 package (17). Three sets of simulations were run: (*i*) with the AMBER ff03* force field (18) for the protein with TIP4P-Ew water (19) in a truncated octahedral cell with an initial size of 65 Å, containing 6,637 water molecules; (*ii*) with the AMBER ff03* force field (18) for the protein with TIP3P water (20) in a truncated octahedral cell with an initial size of 65 Å, containing 6,637 water molecules; (*iii*) with the OPLS/AA-L force field (18) for the protein with TIP3P water (20) in a cubic cell with an initial size of 60 Å, containing 6,648 water molecules. The simulations were propagated using Langevin dynamics with a time step of 2 fs and a friction coefficient of 1 ps^{-1} on all heavy atoms to maintain a constant temperature, and bond lengths were kept fixed using the LINCS constraint algorithm (21). Long-range electrostatics were calculated using particle mesh ewald (PME) with a grid spacing of 1.2 Å and a cutoff of 9 Å. Initial configurations were obtained by solvating an unfolded chain in a box of water and deleting the overlapping water molecules. To obtain the correct size for the simulation cell, the system was equilibrated for 200 ps at 300 K at a constant pressure of 1 atm using a Parrinello–Rahman barostat (22). From the equilibrated configuration, a 1-ns simulation at constant volume was run at 800 K to generate randomized initial coordinates. Replica exchange molecular dynamics at a constant volume was performed using 32 replicas spanning the temperature range from 275 to 352.5 K in increments of 2.5 K, with each replica initiated from a different configuration drawn from the run at 800 K. Exchange attempts were made every 5 ps, and coordinates were saved every 10 ps. Each replica was run for 140, 150, and 100 ns for simulations *i*, *ii*, and *iii*, respectively. Fig. S5 shows the cumulative R_g over the simulations, indicating that most of the relaxation to equilibrium is complete after ≈ 30 ns. The first 30 ns of the simulations thus were discarded to allow for equilibration. A comparison of the radii of gyration and end-to-end distances as a function of temperature is given for all three force fields in Fig. S6. The data for simulation *iii* appear to be less well converged than those for the other two force fields, as suggested by the increase in end-to-end distance for this simulation, while the overall R_g decreases (Fig. S6).

1. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
2. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
3. Kremer W, et al. (2001) Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur J Biochem* 268:2527–2539.
4. Evstafieva AG, et al. (1995) Overproduction in *Escherichia coli*, purification and properties of human prothymosin α. *Eur J Biochem* 231:639–643.
5. Schuler B (2007) Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol Biol* 350:115–138.
6. Van Der Meer BW, Coker G, III, Chen, S-YS (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).
7. Abbate G, Bernini U, Ragozzino E, Somma F (1978) The temperature dependence of the refractive index of water. *J Phys D Appl Phys* 11:1167–1172.
8. Nozaki Y (1972) The preparation of guanidine hydrochloride. *Methods Enzymol* 26:43–50.
9. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
10. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci USA* 102:2754–2759.

11. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Brownian-motion of ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers* 17:11–31.
12. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B* 112:6137–6146.
13. Benninger RK, et al. (2006) Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal Chem* 78:2272–2278.
14. Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.
15. Gast K, Nöppert A, Müller-Frohne M, Zirwer D, Damaschun G (1997) Stopped-flow dynamic light scattering as a method to monitor compaction during protein folding. *Eur Biophys J* 25:211–219.
16. Provencher SW (1982) CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput Phys Commun* 27:229–242.
17. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
18. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487.
19. Horn HW, et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665–9678.
20. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
21. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472.
22. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182–7190.

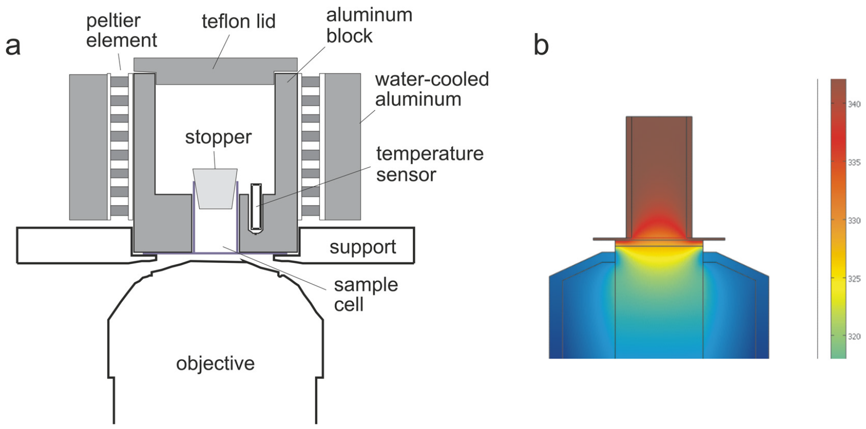


Fig. S1. Temperature-controlled sample holder for confocal single-molecule measurements. (*a*) Schematic of the Peltier-controlled sample holder. Evaporation of immersion water is avoided by a rubber sealing ring between the objective and the cell support. (*b*) The temperature distribution from finite element calculations (objective at room temperature and sample holder at 342 K, prepared using COMSOL Multiphysics; COMSOL) shows a pronounced temperature gradient, illustrating that the temperature must be determined locally (i.e., at the position of the confocal volume just above the cover slide of the sample cell). The calculated temperature at the position of the confocal volume is in good agreement with the values obtained from rhodamine B fluorescence lifetimes.

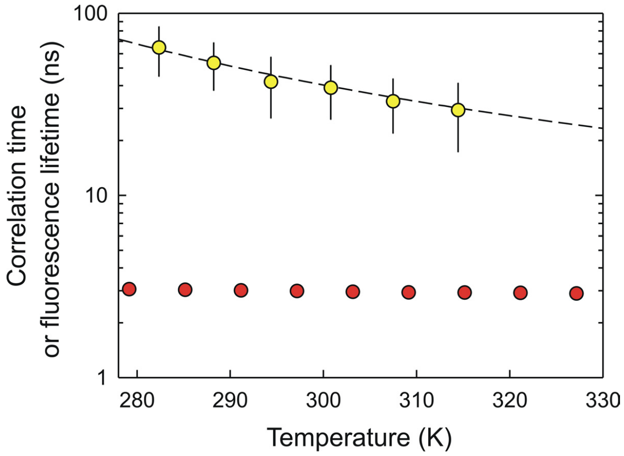


Fig. S2. Temperature dependence of the decay time of the donor fluorescence intensity correlation function (yellow), measured as described in ref. 9, and of the average fluorescence lifetime of the donor (red) of FRET-labeled full-length Csp*Tm* (C2C67) in 5.3 M guanidinium chloride. The dashed line shows a fit to the data assuming that the correlation times scale with the temperature dependence of the viscosity of water. The good agreement indicates that the decrease in solvent viscosity is the dominant cause of faster chain dynamics with increasing temperatures. The error bars indicate our estimate of the experimental uncertainty.

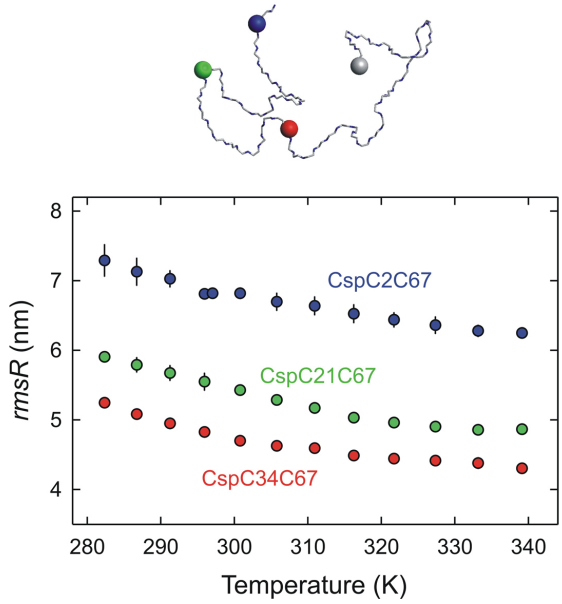


Fig. S3. Temperature-induced unfolded state collapse of Csp*Tm* variants with the FRET pair in different positions within the polypeptide chain in 2.0 M guanidinium chloride. Data acquisition and analysis were performed as in Figs. 1 and 2. Fluorophores were placed at positions 2 and 67 (blue, same variant as in Figs. 1, 2, and 3), 21 and 67 (green), or 34 and 67 (red, illustrated in a schematic of the unfolded chain above the graph). Variants were prepared as described in ref. 2. Error bars indicate standard deviations estimated from at least two independent measurements. The similarity of the changes in intramolecular root-mean-squared distances (*rmsR*) of 14 ± 3% (CspC2C67), 17 ± 2% (CspC21C67), and 18 ± 2% (CspC34C67) provides additional evidence that the temperature-induced compaction of unfolded Csp*Tm* is a global process that affects the entire chain.

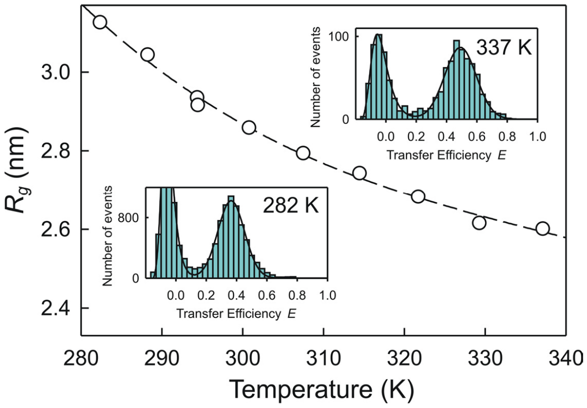


Fig. S4. Temperature-induced collapse of prothymosin α in 0.5 M sodium phosphate buffer. (*Insets*) Corresponding FRET efficiency histograms at 282 and 337 K as examples. An empirical fit used for interpolation is shown as a dashed line.

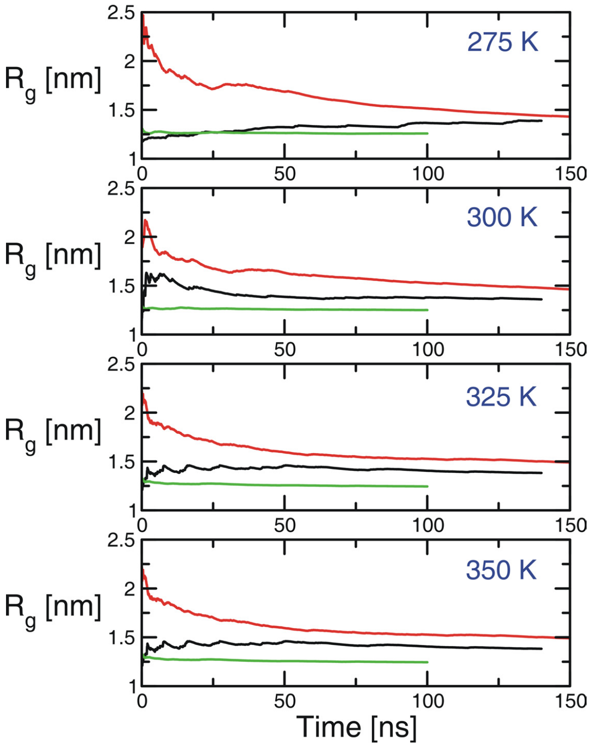


Fig. S5. Cumulative average radius of gyration (R_g) of unfolded Csp*Tm* as a function of temperature for the three sets of simulations: (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green).

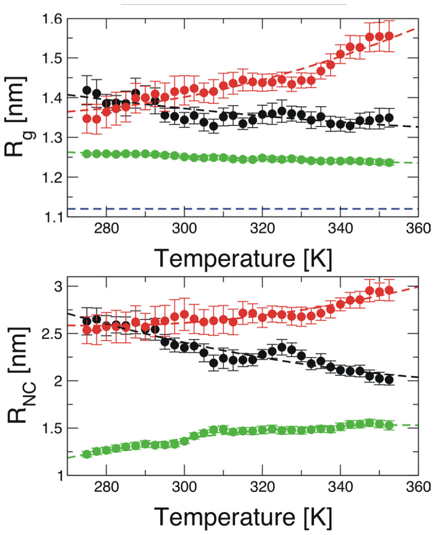


Fig. S6. Radius of gyration (R_g) of unfolded Csp*Tm* (*upper*) as a function of temperature for the three sets of simulations, (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green), and the corresponding mean end-to-end distance (R_{NC}) (*lower*).

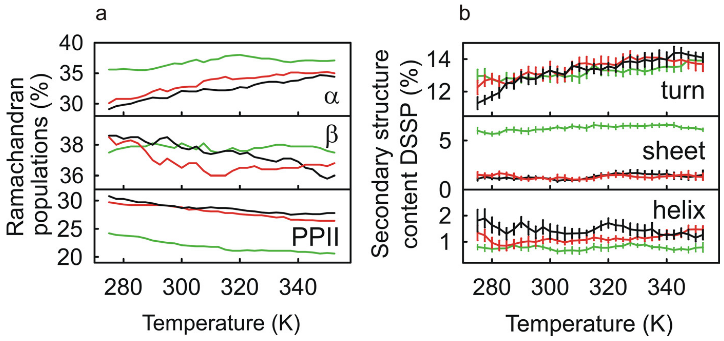


Fig. S7. Populations in a Ramachandran map (*a*) as a function of temperature for the three sets of simulations, (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green), and average secondary structure content from a Database of Secondary Structure of Proteins (DSSP)-based analysis (*b*).

3 Probing Protein–Chaperone Interactions with Single-Molecule Fluorescence Spectroscopy

F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, and B. Schuler.

Contributions of Dominik Hänni:

I chemically modified the GroEL chaperone to allow the immobilization on quartz surfaces. After binding rhodanese to the GroEL chaperones, I immobilized them in a newly developed flow cell and performed TIRF and confocal single-molecule experiments. To automate the data collection on the confocal microscope, I co-developed control software which allowed the autonomous identification and collection of single-molecule fluorescence traces. To analyze the resulting datasets, I implemented automated filtering procedures and image analysis algorithms. This allowed studying the continuous dynamics of GroEL bound rhodanese from nanoseconds to seconds.

Protein Folding

DOI: 10.1002/anie.200800298

Probing Protein–Chaperone Interactions with Single-Molecule Fluorescence Spectroscopy**

*Frank Hillger, Dominik Hänni, Daniel Nettels, Sonja Geister, Michelle Grandin, Marcus Textor, and Benjamin Schuler**

Molecular chaperones are an essential part of the cellular machinery that aids protein folding and assembly in vivo. Particularly remarkable are the members of the Hsp60 class, which encapsulate the folding protein in a central, closed cavity; the most well-studied example is the bacterial GroEL/ES system. Work of the past two decades has resolved many aspects of the processes involved.[1] However, remarkably little is known about the influence of the chaperone on the conformational distributions and folding mechanisms of its substrate proteins.[2] Because of the structural heterogeneity of the nonnative substrate bound to a molecular machine in the 10^6 Da range, its experimental investigation has been difficult with established ensemble methods.[2] Since single-molecule spectroscopy, in particular in combination with Förster resonance energy transfer (FRET), can provide distance and orientational information free of ensemble averaging[3] and allows intramolecular distance dynamics to be observed at equilibrium,[4,5] it is a promising approach to address such questions.[6] Herein, we show how single molecule FRET can be utilized to investigate the nonnative conformation and dynamics of bovine rhodanese, a classic chaperone substrate protein,[7,8] upon interaction with GroEL.

To obtain a transfer efficiency signature suitable for discriminating native and nonnative conformations, two rhodanese variants with complementary donor and acceptor positions (Figure 1) were investigated. Figure 1c–j shows the transfer efficiency histograms determined from photon bursts originating from individual labeled rhodanese molecules freely diffusing through the observation volume of the confocal instrument. As expected, the rhodanese variant with the labels at the domain interface (Figure 1a) shows a mean transfer efficiency $\langle E\rangle$ close to 1 in its native state (Figure 1c); for the variant with labels at the ends of the interdomain linker (Figure 1b), $\langle E\rangle = 0.69$ (Figure 1d), which corresponds to a distance of 4.7 nm, in good agreement with the distance of 4.5 nm in the crystal structure.[9] In the unfolded state at 5 M guanidinium chloride (GdmCl), $\langle E\rangle$ scales with the sequence separation of the labeling sites (Figure 1e,f), as expected.

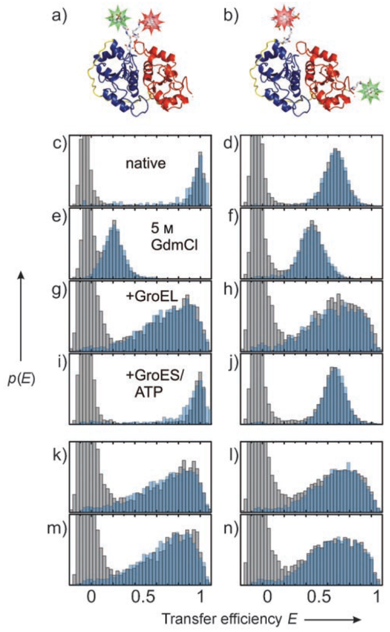


Figure 1. Native structures (based on PDB ID 1RHD) and transfer efficiency (*E*) histograms of rhodanese variants D102C–D219C (interface variant, panels on the left) and K135C–K174C (linker variant, panels on the right). a,b) Alexa Fluor488 and Alexa Fluor594 were coupled to the cysteine residues introduced by site-directed mutagenesis. c–j) *E* histograms of rhodanese variants under native (c,d) and denaturing conditions (5 M GdmCl, e,f), bound to GroEL upon dilution from GdmCl (g,h), and after refolding by addition of GroES/ATP (i,j), bound to GroEL after incubation with folded rhodanese at 30 °C for 16 h (k,l), and after dilution from 0.1 M phosphoric acid (m,n).The gray histograms were recorded with donor excitation only. For the blue histograms, pulsed interleaved excitation[14] was used. *p*(*E*) = relative event frequency.

---

[*] Dr. F. Hillger, D. Hänni, Dr. D. Nettels, S. Geister, Prof. Dr. B. Schuler
Universität Zürich, Biochemisches Institut
8057 Zürich (Switzerland)
Fax: (+41) 44-635-5907
E-mail: schuler@bioc.uzh.ch
Homepage: http://www.bioc.uzh.ch/schuler
Dr. M. Grandin, Prof. Dr. M. Textor
ETH Zürich, Departement Materialwissenschaft
8093 Zürich (Switzerland)

[**] We thank A. Szabo for discussions and advice on anisotropy simulations, A. Plückthun and M. Kawe for discussions and samples of GroEL, H. Hofmann and D. Streich for discussions, the late P. Horowitz for a plasmid encoding rhodanese, and G. Lorimer for a plasmid encoding SR1. This work was supported by the VolkswagenStiftung, the Schweizerische Nationalfonds, the Swiss National Center of Competence in Research for Structural Biology, and the Human Frontier Science Program.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200800298.

Upon dilution of labeled rhodanese unfolded in GdmCl into buffer containing an excess of unlabeled GroEL, rhodanese becomes bound to the chaperone quantitatively, as evidenced by analytical size-exclusion chromatography (data not shown). We exclude the possibility of substrate protein binding to both chaperone rings by using the single-ring variant of GroEL, SR1,which binds to the substrate in a 1:1 complex.[10,11] Experiments with wild-type tetradecameric GroEL gave results essentially identical to the ones presented here. The transfer efficiency histograms of SR1-bound rhodanese (Figure 1 g,h) exhibit a pronounced broadening, indicating the presence of static heterogeneity on the observation time scale (≈ 1 ms, duration of a fluorescence burst). For a random conformational distribution, we would expect a transfer efficiency that scales with the sequence separation of the dyes, as in the denaturant-unfolded state (Figure 1 e,f). In contrast, we observe maxima of the transfer efficiency histograms close to the values found in the native state (Figure 1 c,d), suggestive of a bias towards the native topology for rhodanese bound by the chaperone. The presence of very low intramolecular transfer efficiencies that could be hidden under the "donor only" peak[12] at $E \approx 0$ was excluded in experiments using alternating excitation of donor and acceptor[13,14] (Figure 1 c–n). The slight but reproducible difference in shape between the transfer efficiency histograms of the two chaperone-bound rhodanese variants (Figure 1 g,h) suggests that the E histograms provide a characteristic signature for the conformation of the substrate protein. Remarkably, the shapes of the transfer efficiency histograms are independent of how rhodanese is denatured (Figure 1 g,h,k–n), implying that the chaperone-bound conformation does not reflect the conformational distribution under unfolding conditions, but rather resembles a folding intermediate that is formed rapidly upon dilution into the SR1 solution (Figure 1 g,h,m,n), and that is also accessible from the native state under mildly destabilizing conditions (Figure 1 k,l).[15] After addition of ATP and the cochaperone GroES to the rhodanese–GroEL complex, the E distributions characteristic of the native structures are recovered[*] (Figure 1 c,d,i,j), demonstrating that labeled rhodanese is a fully functional chaperone substrate.

To probe the dynamics of the rhodanese–chaperone complex, we used correlation experiments employing a Hanbury Brown and Twiss setup.[4] Figure 2 a shows that rhodanese unfolded in 5 M GdmCl exhibits rapid intramolecular chain dynamics on a time scale of ≈ 70 ns. This time scale is very similar to that observed for the unfolded cold shock protein Csp*Tm*[4] and the Sup35 NM domain.[**][16] How do the dynamics of the denatured state change upon association with GroEL? The same measurement on rhodanese bound to GroEL yields a correlation with a decay of 0.2 μs (Figure 2 b), which at first sight could be misinterpreted as slowed distance dynamics. But the pronounced sensitivity of the correlation amplitude on the directions of polarization that are correlated (Figure S1 in the Supporting Information) indicates a strong contribution from rotational motion of the entire GroEL–rhodanese complex, which occurs exactly on this time scale.[*][18] To quantify the relative contributions of rotational and distance dynamics, we compare GroEL-bound rhodanese labeled with a FRET pair to GroEL-bound rhodanese labeled only with a donor chromophore. As shown in Figure 2 b, the two samples exhibit the same decay time of the correlation function. The ratio of the two correlations does not indicate the presence of an additional component (Figure 2 c), suggesting that the observed correlation is entirely due to rotation, and that distance dynamics are absent on this time scale. Additional evidence for the lack of distance dynamics comes from the pronounced intensity correlations of polarized acceptor emission upon donor excitation, which exhibit the same 0.2 μs decay (Figure S2 in the Supporting Information). This result shows that the relative orientation of donor and acceptor is rather invariant on this time scale, arguing that the same is true for their distance. The nanosecond chain

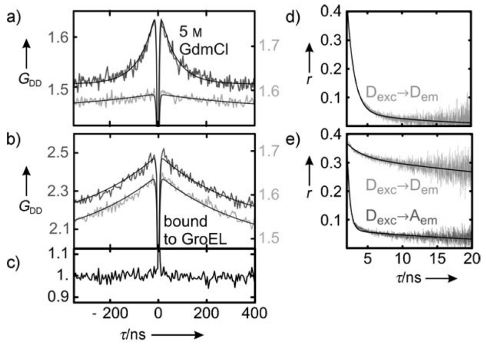


Figure 2. Dynamics of rhodanese. a–c) Donor–donor fluorescence intensity autocorrelation functions G_{DD} from Hanbury Brown and Twiss start–stop experiments.[4] Correlation functions are shown for the linker variant unfolded in 5 M GdmCl (a) and bound to GroEL (b). Dark gray lines show the correlation functions for the FRET-labeled, light gray lines for the donor-only-labeled rhodanese K174C. The black curves in (a) and (b) show fits to the correlation data including photon antibunching.[4] c) The normalized ratio of the two correlation functions from (b) indicates the absence of distance dynamics. d,e) Anisotropy decays for donor-only-labeled rhodanese K174C (light gray) under denaturing conditions (d) and bound to GroEL (e). The dark gray data in (e) show the fluorescence anisotropy decay of the acceptor upon excitation of the donor for the linker variant bound to GroEL. The black curves in (d) and (e) represent fits to Equation (1).

---

[*] Under our conditions, folded rhodanese is not confined within the cage.

[**] Note, however, that (in contrast to Csp*Tm*[4]), even singly labeled rhodanese exhibits some bunching on this time scale, albeit with lower amplitude (Figure 2 a), which complicates a quantitative analysis. This behavior is similar to recent observations for a Sup35 fragment, which were attributed to quenching of the fluorophores by aromatic residues in the chain.[16]

[*] In contrast to the magic angle configuration possible in conventional fluorimeters, the geometry of confocal epifluorescence instruments complicates the elimination of polarization effects on the correlation functions.[17]

dynamics observed in denaturant-unfolded rhodanese are thus suppressed when the protein is bound to the chaperone.

To investigate the presence of distance dynamics on longer time scales, we first employed subpopulation-specific fluorescence correlation spectroscopy on freely diffusing rhodanese–GroEL complexes (Figure 3). We correlated only signal from FRET-labeled molecules with $E > 0.2$ to minimize the contribution from donor-only labeled species, and we used donor–acceptor cross-correlation analysis to minimize the contribution of microsecond triplet dynamics.[19] Distance fluctuations would then result in an anticorrelated signal, that is, a rise in the correlation function. However, the correlation curves show no evidence for the presence of distance fluctuations up to ≈100 µs. To extend the accessible time scales beyond the diffusion time through the confocal volume, SR1–rhodanese complexes were immobilized on cover slides coated with biotinylated poly(L-lysine)-*graft*-poly(ethylene glycol) (PLL-*g*-PEG).[20] Individual complexes on the surface were identified by sample scanning and were observed individually for several seconds until the chromophores bleached. Surprisingly, donor–acceptor cross-correlation analysis of these data indicates the absence of long-range distance dynamics in chaperone-bound rhodanese even on long time scales. The decay of the correlation function setting in at times > 10 ms is caused by irreversible photobleaching, as indicated by the divergence of donor–acceptor and acceptor–donor cross-correlations (Figure 3).[21] The absence of large-amplitude distance fluctuations is also supported by the large width of transfer efficiency histograms from different observation or binning times (Figure 3, insets), indicating the presence of static heterogeneity on time scales up to at least 100 ms.

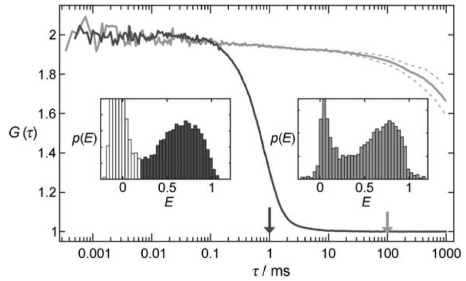


Figure 3. Normalized fluorescence intensity donor–acceptor cross-correlations and transfer efficiency histograms of the rhodanese-linker variant bound to GroEL (free diffusion: dark gray, surface-immobilized: light gray; corresponding binning times for *E* histograms indicated by arrows). Dashed lines indicate the individual cross-correlations (D→A and A→D, shown for $\tau > 0.5$ ms). For freely diffusing molecules, only events with $E > 0.2$ were used for the correlation (dark gray).

Finally, we need to establish the structural origin of the large width of the transfer efficiency distributions of chaperone-bound rhodanese (Figure 1). Static heterogeneity of the transfer rate can originate from either a distribution of intramolecular distances or a distribution of donor–acceptor orientations.[*] For rhodanese singly labeled with donor or acceptor and unfolded in 5 M GdmCl, the anisotropy decay $r(t)$ is dominated by a single component with a time constant of ≈1 ns (Figure 2d), indicating rapid and complete reorientation of the dyes, and thus justifying the common approximation of $\kappa^2 \approx 2/3$ for the orientational factor in Förster theory.[22] Upon binding to SR1, however, the anisotropy of all singly labeled variants (D102C, K135C, K174C, D219C) increases drastically, and the majority of the anisotropy decay occurs on the time scale of rotation of the entire rhodanese–SR1 complex (> 100 ns, Figure 2e). Consequently, the orientational restriction of the dyes must be taken into account to obtain distance information.

To this end, we analyze the fluorescence anisotropy decays of our singly labeled rhodanese variants with Equation (1), which describes the decay as the combined effect of restricted dye rotation (τ_{eff}) and the rotational motion of the entire protein–chaperone complex (τ_{M}).[23]

$$r(t) = \left((r_0 - r_\infty)e^{-t/\tau_{\text{eff}}} + r_\infty\right)e^{-t/\tau_{\text{eff}}} \qquad (1)$$

Here, r_0 is the limiting anisotropy of the dyes,[**] and r_∞ is the residual anisotropy assuming no rotation of the macromolecule carrying the dye. Assuming restricted angular diffusion in a cone as the simplest plausible model for the motion of the chromophores[24] (Figure 4a), the semiangle Θ_{max} of the cone can be calculated from Equation (2),[23,25] yielding for all our variants and dyes values between 17° and 19°.[***]

$$r_\infty = r_0\left(\frac{1}{2}\cos\Theta_{\text{max}}(1 + \cos\Theta_{\text{max}})\right)^2 \qquad (2)$$

Important additional information about the relative orientation of the dyes comes from the anisotropy decay of the acceptor upon donor excitation (Figure 2e): in this case, the residual anisotropy approaches zero for both chaperone-bound variants, indicating an angular distribution of the cone axes that is close to random. An alternative explanation, a narrow relative orientation close to the magic angle of 54.7°, can be excluded, because this would result in an apparent fundamental anisotropy of zero for the acceptor anisotropy decay upon donor excitation, which is incompatible with our obervations (Figure 2e). Additionally, a narrow distribution

---

[*] Heterogeneity in the quantum yields of the dyes originating, for example, from differences in the local environment can be excluded because of the agreement of fluorescence lifetimes (both in ensemble and single-molecule measurements) of the acceptor in FRET-labeled and the donor in singly labeled rhodanese on GroEL, respectively, with the lifetimes of the dyes on protein unfolded in 5 M GdmCl.

[**] $r_0 = 0.38$ was determined in a matrix of 99% glycerol at −10°C.

[***] The lack of binding to GroEL of free dyes and several other small proteins and peptides labeled with the same dyes (size exclusion chromatography data not shown) indicates that the interaction of rhodanese with GroEL is dominated by the polypeptide and that the orientational restriction of the dyes results largely from steric constraints in the rhodanese–chaperone complex.

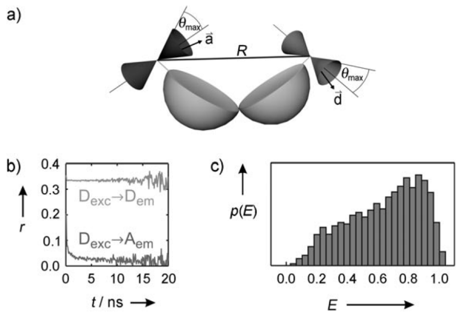


Figure 4. Simulations and structural interpretation. a) Assuming angular diffusion of the fluorophore dipoles $\vec{d}$ and $\vec{a}$ in cones of half angle Θ_{max} on the protein surface, and a narrow Gaussian distribution of distance R, we can account for both the characteristic anisotropy decays[34] (b) and the broad transfer efficiency histograms (c) observed experimentally (cf. Figures 1 and 2e). The two domains of rhodanese are indicated by the gray hemispheres in (a).

of relative orientations would at the same time require a broad distribution of distances to account for the broad transfer efficiency histograms we observe (Figure 1g,h,k–n), but this combination is physically implausible.

To interpret the experimental results quantitatively, we thus simulated the transfer process between orientationally restricted dipoles based on the simplest plausible model for our system (Figure 4): we assume that the relative orientation of the cones is fixed for every individual rhodanese–GroEL complex but randomly distributed from molecule to molecule. This assumption leads to anisotropy decays (Figure 4b) very similar to the experimental ones (Figure 2e), even for the characteristic decay of the acceptor anisotropy upon donor excitation. If we now assume a normal distribution of interdye distances R (Figure 4a) and adjust its mean and standard deviation to maximize the agreement between simulated and observed transfer efficiency histograms (Figure 4c and Figure S3 in the Supporting Information), we obtain distance distributions for GroEL-bound rhodanese with a mean distance of $\approx 4.5 \pm 0.5$ nm and a width of $\approx 0.5 \pm 0.2$ nm for both variants. For the linker variant, this is approximately the same value as in the native structure, but for the interface variant it is significantly larger, suggesting a large separation of the two rhodanese domains. If we used the mean distances for our two variants as constraints to adjust the relative orientation of the native domains, we would obtain a rhodanese conformation that is highly suggestive of binding to the rim of the GroEL ring, which is lined by hydrophobic residues that act as binding sites,[26] an arrangement that is in accord with a number of previous results.[27–29]

In summary, we have used a novel analysis combining time-resolved fluorescence anisotropy decays, single-molecule FRET experiments, and simulations to obtain quantitative information from a system with orientationally restricted chromophores, a situation that has been observed repeatedly in FRET experiments involving protein–chaperone interactions.[6,30] In our analyis, the donor and acceptor anisotropy decays define the opening angles of the cones constraining fluorophore rotation; the acceptor decay upon donor excitation constraines the relative orientational distribution of the cones; and the shape of the transfer efficiency histograms then define the mean and width of the distance distributions. What emerges from these measurements, together with the long-range dynamic information available from subpopulation-specific correlation functions, is the picture of a rather well-defined ensemble of rhodanese conformations that resembles a partially structured folding intermediate when bound to the chaperone GroEL. Interestingly, the lack of long-range distance dynamics does not seem to preclude local structural fluctuations evident from protease susceptibility,[27,28] or NMR[18,31] or fluorescence spectroscopy.[8,27] Our results illustrate the potential for extracting quantitative structural information from FRET experiments even in cases where large anisotropies demand that orientational effects be taken into account, and provide an important step towards investigating the role of cellular factors in protein folding.

Experimental Section

Proteins were prepared as described previously.[10,32] Binding of rhodanese to SR1 was achieved as follows: A) Rhodanese unfolded in 5 M GdmCl was rapidly diluted tenfold into folding buffer (0.1 M potassium phosphate, 5 mM magnesium chloride, 200 mM 2-mercaptoethanol, 0.001 % Tween 20, 1 mM EDTA, pH 7.0) containing at least a tenfold molar excess of SR1 heptamers. B) Same as (A), but unfolding was carried out in 0.1 M phosphoric acid. C) Rhodanese was incubated at 30 °C for 16 h in folding buffer with a tenfold molar excess of SR1 heptamers. Complete binding was assessed on a TSK 5000 PWXL column (TOSOH Bioscience) with fluorescence detection.

For surface immobilization, SR1 was biotinylated using (+)-biotin N-hydroxysuccinimide ester in a molar ratio of 1:7. A solution of 0.1 mg mL^{-1} PLL(20)-g[3.5]-PEG(2)/PEG(3.4)-biotin (50 %)[20] in 10 mM potassium phosphate, pH 7.0, was applied to a custom-made quartz flow cell. After 15 min of incubation, the flow cell was washed with 0.1 M potassium phosphate, 5 mM magnesium chloride, 1 mM EDTA, pH 7.0; then 1 mg mL^{-1} avidin was applied in the same buffer. After 15 min of incubation, the flow cell was washed thoroughly, and 250–500 nM GroEL–rhodanese preparation was applied. The cell was incubated for 5 min and then washed with buffer.

Anisotropy decay data were recorded with a custom-built fluorescence lifetime spectrometer using 1 μM samples of labeled protein. Single-molecule FRET measurements were performed as previously described[4,32,33] using an adapted MicroTime 200 confocal microscope (PicoQuant, Berlin). For additional details, see the Supporting Information.

Received: January 21, 2008
Revised: April 29, 2008

Keywords: chaperones · correlation spectroscopy · FRET · protein folding · single-molecule studies

---

[1] D. Thirumalai, G. H. Lorimer, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 245–269; H. Grallert, J. Buchner, *J. Struct. Biol.* **2001**, *135*, 95–103; J. C. Young, V. R. Agashe, K. Siegers, F. U. Hartl, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 781–791; A. L. Horwich, G. W. Farr, W. A. Fenton, *Chem. Rev.* **2006**, *106*, 1917–1930.

[2] W. A. Fenton, A. L. Horwich, *Q. Rev. Biophys.* **2003**, *36*, 229–256.

[3] G. Haran, *J. Phys. Condens. Matter* **2003**, *15*, R1291–R1317; B. Schuler, *ChemPhysChem* **2005**, *6*, 1206–1220; X. Michalet, S. Weiss, M. Jäger, *Chem. Rev.* **2006**, *106*, 1785–1813; B. Schuler, W. A. Eaton, *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

[4] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655–2660.

[5] E. V. Kuzmenkina, C. D. Heyes, G. U. Nienhaus, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15471–15476.

[6] S. Sharma, K. Chakraborty, B. K. Müller, N. Astola, Y. C. Tang, D. C. Lamb, M. Hayer-Hartl, F. U. Hartl, *Cell* **2008**, *133*, 142–153.

[7] J. A. Mendoza, E. Rogers, G. H. Lorimer, P. M. Horowitz, *J. Biol. Chem.* **1991**, *266*, 13044–13049.

[8] J. Martin, T. Langer, R. Boteva, A. Schramel, A. L. Horwich, F. U. Hartl, *Nature* **1991**, *352*, 36–42.

[9] J. H. Ploegman, G. Drent, K. H. Kalk, W. G. J. Hol, *J. Mol. Biol.* **1978**, *123*, 557–594.

[10] A. L. Horwich, S. G. Burston, H. S. Rye, J. S. Weissman, W. A. Fenton, *Methods Enzymol.* **1998**, *290*, 141–146.

[11] J. S. Weissman, H. S. Rye, W. A. Fenton, J. M. Beechem, A. L. Horwich, *Cell* **1996**, *84*, 481–490.

[12] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2754–2759.

[13] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8936–8941.

[14] B. K. Müller, E. Zaychikov, C. Bräuchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522.

[15] T. Shibatani, G. Kramer, B. Hardesty, P. M. Horowitz, *J. Biol. Chem.* **1999**, *274*, 33795–33799; K. E. Smith, P. A. Voziyan, M. T. Fisher, *J. Biol. Chem.* **1998**, *273*, 28677–28681.

[16] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2649–2654.

[17] Ü. Mets in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer, Berlin, **2001**.

[18] R. Horst, E. B. Bertelsen, J. Fiaux, G. Wider, A. L. Horwich, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12748–12753.

[19] P. Schwille in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer, Berlin, **2001**; C. Eggeling, P. Kask, D. Winkler, S. Jäger, *Biophys. J.* **2005**, *89*, 605–618.

[20] N. P. Huang, J. Vörös, S. M. De Paul, M. Textor, N. D. Spencer, *Langmuir* **2002**, *18*, 220–230.

[21] C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, C. A. M. Seidel, *J. Phys. Chem. A* **2006**, *110*, 2979–2995.

[22] B. W. Van Der Meer, G. Coker III, S. Y. S. Chen, *Resonance Energy Transfer: Theory and Data*, VCH, New York, **1994**.

[23] G. Lipari, A. Szabo, *Biophys. J.* **1980**, *30*, 489–506.

[24] G. F. Schröder, U. Alexiev, H. Grubmüller, *Biophys. J.* **2005**, *89*, 3757–3770.

[25] K. Kinosita, S. Kawato, A. Ikegami, *Biophys. J.* **1977**, *20*, 289–305.

[26] K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, P. B. Sigler, *Nature* **1994**, *371*, 578–586.

[27] J. A. Mendoza, M. C. Butler, P. M. Horowitz, *J. Biol. Chem.* **1992**, *267*, 24648–24654.

[28] R. Hlodan, P. Tempst, F. U. Hartl, *Nat. Struct. Biol.* **1995**, *2*, 587–595.

[29] P. Thiyagarajan, S. J. Henderson, A. Joachimiak, *Structure* **1996**, *4*, 79–88; A. van der Vaart, J. P. Ma, M. Karplus, *Biophys. J.* **2004**, *87*, 562–573; G. W. Farr, K. Furtak, M. B. Rowland, N. A. Ranson, H. R. Saibil, T. Kirchhausen, A. L. Horwich, *Cell* **2000**, *100*, 561–573.

[30] Z. Lin, H. S. Rye, *Mol. Cell* **2004**, *16*, 23–34.

[31] R. Zahn, S. Perrett, G. Stenberg, A. R. Fersht, *Science* **1996**, *271*, 642–645; R. Zahn, C. Spitzfaden, M. Ottiger, K. Wüthrich, A. Plückthun, *Nature* **1994**, *368*, 261–265.

[32] F. Hillger, D. Nettels, S. Dorsch, B. Schuler, *J. Fluoresc.* **2007**, *17*, 759–765.

[33] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 105–110.

[34] Rotation of the rhodanese-GroEL complex was not included in the simulations, which therefore lack the slow relaxation component ($\tau_M = \infty$ in Eq. 1).

Supporting information

Probing protein-chaperone interactions with single molecule fluorescence spectroscopy

Frank Hillger, Dominik Hänni, Daniel Nettels, Sonja Geister, Michelle Grandin, Marcus Textor, and Benjamin Schuler

Materials and Methods

Protein Preparation and Labeling. Cysteine variants of rhodanese were prepared and labeled with Alexa 488 and Alexa 594 (R_0 = 5.4 nm) as described previously[1]. Cysteine residues were introduced by site-directed mutagenesis at positions 135 and 174 ("linker variant") and 102 and 219 ("interface variant"), respectively. After labeling, thiosulfate was removed from the active site of rhodanese by reaction with potassium cyanide and subsequent size exclusion chromatography. GroEL-SR1 was expressed and purified essentially as described by Horwich *et al.*[2] and stored at -80 °C as a suspension of precipitate in 2.7 M ammonium sulfate.

Preparation of Rhodanese-Chaperone Complexes. SR1 ammonium sulfate precipitate was resolubilized at 10 to 20 µM SR1 (heptamer) in folding buffer (0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 200 mM 2-mercaptoethanol (Fluka Ultra), 0.001 % Tween 20 (Pierce), 1 mM EDTA (Biosolve), pH 7.0). Binding of rhodanese to SR1 was achieved as follows: **A**) Rhodanese unfolded in 5 M GdmCl (Pierce), 0.1 M potassium phosphate, 200 mM 2-mercaptoethanol, 0.001 % Tween 20, 1 mM EDTA, pH 7.0 was rapidly diluted tenfold into folding buffer containing at least a tenfold molar excess of SR1 heptamers. **B**) As **A,** but unfolding was carried out by tenfold dilution of native rhodanese into 0.1 M phosphoric acid. **C**) Rhodanese was incubated at 30 °C over night in folding buffer containing a 10-fold molar excess of SR1 heptamers. Complete binding was assessed by analytical size exclusion chromatography on a TSK 5000 PWXL (TOSOH Bioscience) and fluorescence detection. In all cases, rhodanese was bound completely to SR1. Single-molecule experiments and size exclusion chromatography indicated the absence of unbound rhodanese molecules or aggregates[1].

Note that rhodanese shows a pronounced collapse upon dilution from denaturant within the dead time of ensemble stopped-flow mixing experiments, indicating the formation of a compact structural ensemble before the slower phases of folding ensue. The dominant phase of refolding for our labeled rhodanese variants occurs with half times of ~15 min, within the range of reactivation times of unlabeled rhodanese reported in the literature[3], suggesting that labeling does not strongly interfere with the folding mechanism. The widths of the transfer efficiency histograms (Fig. 1 main text) for native, unfolded, and refolded rhodanese interface variant were 0.08±0.01, 0.08±0.01, and 0.10±0.01, respectively, and for the linker variant 0.11±0.01, 0.10±0.01, and 0.10±0.01, respectively. At the exceedingly low protein concentrations used in our experiments, aggregation rates become negligible and the native state is reached reversibly upon refolding also in the absence of GroEL/ES.

Biotinylation and Surface Immobilization of GroEL/SR1. Ammonium sulfate-precipitated SR1 was transferred to 100 mM $NaHCO_3$ (Roth) using a 5 ml HiTrap desalting column (GE Amersham), followed by concentration to 30-40 mg/ml in Centricon YM-3 filter devices (Amicon). (+)-Biotin N-hydroxy-succinimide ester (Sigma) was dissolved in dry DMSO (Pierce) and added to concentrated SR1 in a molar ratio of 1 to 7. After incubation for 4 hours, the reaction mixture was purified and transferred to 0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 1 mM EDTA (Biosolve), pH 7.0, using a Superdex 200 10/300 GL (Amersham) gel filtration column. Biotinylation was analyzed using ESI TOF mass spectroscopy and samples were stored at 5 mg/ml at -80 °C. 0.1 mg/ml PLL(20)-g[3.5]-PEG(2)/PEG(3.4)-Biotin (50%)[4] was dissolved in 10 mM potassium phosphate (Fluka Ultra), pH 7.0, and applied to a custom-made quartz flow cell. After 15 min of incubation, the flow cell was washed with 0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 1 mM EDTA (Biosolve), pH 7.0, then 1 mg/ml Avidin D (Vector) were applied in the same buffer. After 15 min of incubation, the flow cell was washed vigorously. The GroEL/rhodanese complex was prepared as described above and separated from monomeric rhodanese and rhodanese aggregates by size exclusion chromatography. 250-500 nM of GroEL/rhodanese preparation were applied to the flow cell and incubated for 5 min before the cell was washed with buffer.

Single Molecule Fluorescence Spectroscopy. Observations of single-molecule fluorescence were made using a MicroTime 200 confocal single molecule instrument (PicoQuant, Berlin, Germany) equipped with a Olympus UplanApo 60x/1.20W objective and four detection channels. Donor excitation pulses of 470 nm wavelength are generated by a diode laser (LDH 470, PicoQuant). Except for PIE measurements (see below) the pulse repetition rate was 40 MHz, resulting in an average power of 100 µW entering the completely illuminated back aperture of the objective. Laser light and fluorescence emission are separated using a dual band beam splitter (z488/568, Chroma), followed by a 100 µm pinhole. Sample fluorescence was further separated by a polarizing beam splitter cube into components parallel and perpendicular polarized with respect to the excitation light. Subsequently, both components were further divided into donor and acceptor photons by means of dichroic mirrors (585DCXR, Chroma), filtered (donor emission filters: Chroma HQ525/50, Omega 525AF45, acceptor emission filters: Chroma HQ650/100 and HQ600LP), and focused on avalanche photodiodes (PerkinElmer Optoelectronics SPCM-AQR-15). A TimeHarp 200 counting card (PicoQuant) records the arrival time of every detected photon with a time resolution of 100 ns. In addition, the time separation between excitation pulse and photon detection is stored with 38 ps resolution. For pulsed interleaved excitation[5] PIE, the wavelength range used for acceptor excitation was selected with a z582/15 band pass filter (Chroma) from the emission of an SC-450-4 supercontinuum fiber laser (Fianium, UK) driven at 15 MHz. The emission is monitored by a fast photodiode (DET10A/M, Thorlabs), which triggers the LDH 470 driver to deliver single laser pulses. The time separation between the acceptor and donor excitation pulses of 33 ns was adjusted with a delay line. FCS data down to 0.1 µs lag time (Fig. 3) were recorded and calculated using the software SymphoTime (PicoQuant). Fluorescence correlation data in the sub-microsecond range (down to 256 ps lag time, Fig. 2) were recorded and analyzed as described previously using a Hanbury Brown and Twiss detection scheme[6], a continuous wave solid-state diode-pumped laser (Coherent Sapphire 488-200) operating at 488 nm (average radiant power at the sample: 100µW) and a PicoHarp 300 or a HydraHarp 400 photon counting module (PicoQuant). Subpopulation-specific donor-acceptor cross-correlations were obtained by correlating only photons from bursts assigned to the unfolded subpopulation.

For experiments on surface-immobilized molecules, the MicroTime 200 (PicoQuant) was used in combination with a 488 nm cw laser. Sample scanning with an x/y-piezo-stage (P-733 and E-710, Physik Instrumente) in combination with custom-developed software allowed the automated identification and data acquisition of immobilized molecules. In area scans over regions of $80 \times 80 \mu\text{m}^2$ (0.156 nm point resolution, 5 µW average radiant power at the sample), the locations of the individual molecules were identified and measured for five seconds each. For the data shown in Fig. 2 of the main text, 1297 trajectories were analyzed. To quantify the background intensity, 30 trajectories were measured at positions lacking labeled protein.

Single Molecule Data Reduction and Analysis. Successive photons detected in any channel and separated by less than 100 µs were combined into one burst. A burst was retained as a significant event if the total number of counts exceeded 50 (after correction for background, differences in quantum yields, the different collection efficiencies of the detection channels, cross-talk, and direct excitation of the acceptor)[7]. In addition, bursts with a significant likelihood for acceptor photobleaching during the burst were discarded (see below). Transfer efficiencies were calculated as $E = n_A / (n_D + n_A)$, where n_D and n_A are the corrected donor and acceptor counts of each burst.

For calculating correlation functions from measurements on immobilized molecules, the trajectories were merged and donor and acceptor intensities cross-correlated. For normalization, each correlation was divided by the mean fluorescence calculated from the first 10 data points of each correlation. For transfer efficiency histograms, the trajectory data were binned in 100 ms intervals and corrected for background. For each trajectory the maximum change in the sum of donor and acceptor fluorescence signal was calculated and only trajectories were selected were this difference was >200 counts and did not exceed 1200 counts. The trajectories were truncated at 1 s to minimize the contribution of background. Transfer efficiency histograms were calculated for each bin of the trajectories using $E = n_A / (n_A + \gamma n_D)$, where $\gamma = 1.2$ corrects for the difference in quantum yield of donor and acceptor dyes as well as the detection efficiencies of the channels.

Identification of Fluorescence Bursts Affected by Photobleaching. Assume that during a burst of duration T, n_D donor photons and n_A acceptor photons were detected at times $t_{D,1} \ldots t_{D,n_D}$ and $t_{A,1} \ldots t_{A,n_A}$, respectively. With the average values $\overline{t}_D = 1/n_D \sum_i t_{D,i}$ and $\overline{t}_A = 1/n_A \sum_i t_{A,i}$, we define the burst asymmetry as $\alpha_{DA} = \overline{t}_D - \overline{t}_A$. In the limit of $n_D, n_A \to \infty$, the asymmetry is zero if donor and acceptor dyes emit photons continuously with rates, $\tau_{D,A}^{-1} = n_{D,A} / T$, during the complete burst duration T. However, if the acceptor dye bleaches at a certain time during the photon burst, the acceptor photon rate will drop to the background level, and the donor rate will increase due to the interrupted energy transfer, resulting in a burst asymmetry $\alpha_{DA} > 0$. For finite photon numbers n_D and n_A, shot noise of the calculated α_{DA} has to be taken into account. Here we calculate the standard deviation of α_{DA} for the ideal case that no photobleaching occurs during the burst. For a randomly chosen burst photon i, the detection time t_i is then uniformly distributed, with a constant probability density function $p(t_i) = 1/T$, such that $\int_0^T p(t_i)dt_i = 1$. The expectation value for t_i is then $\bar{t}_i = \int_0^T t_i p(t_i)dt_i = T/2$ and the variance is $\sigma_i^2 = \int_0^T (t_i - \bar{t}_i)^2 p(t_i)dt_i = T^2/12$. With these relations, the variance of $\bar{t}_D$ and $\bar{t}_A$ defined above result as $\Delta\bar{t}_{D,A}{}^2 = \sum_i \sigma_i^2 / n_{D,A}^2 = T^2 / 12 n_{D,A}$. Finally, we obtain for the standard deviation of α_{DA}:

$$\sigma_{DA} = \frac{T}{2\sqrt{3}} \left(\frac{1}{n_D} + \frac{1}{n_A} \right)^{1/2} . \qquad (1)$$

This approximation is valid if the dead time of the detectors (≈ 100 ns) is short compared to the inter-photon times $\tau_{D,A}$ during the bursts (on average ~10 µs). To exclude bursts with a significant deviation of α_{DA} from zero, indicating that a bleaching process is likely to have occurred, bursts with $|\alpha_{DA}| > \sigma_{DA}$ were not included in the analysis. An example of the analysis is shown in Fig. S4.

Time-Resolved Anisotropy. Anisotropy decay data were recorded with a custom-built fluorescence lifetime spectrometer using samples of 1 µM labeled protein. The fluorophores were excited at 470 nm by vertically polarized laser pulses with 20 MHz repetition rate. Collimated fluorescence light, emitted in perpendicular direction with respect to the laser beam, passed an analyzing Glan-Thompson polarizer and a band pass filter (525AF45, Omega Filters, or HQ640/100, Chroma) before it was focused onto a microchannel plate photomultiplier tube (PMT R3809U-50, Hamamatsu). The output pulses of the PMT were pre-amplified (PAM 102-M, PicoQuant), and times between laser pulses and fluorescence photon arrival times were recorded by a photon counting module (PicoHarp 300, PicoQuant). Fluorescence decay histograms with 8 ps bin width were formed. The overall instrument response function (IRF) of the system was measured to have a FWHM of $\Delta_{IRF} = 80$ ps. Two fluorescence decays $I_V(t)$ and $I_H(t)$ were measured for each sample with the analyzing polarizer in vertical and in horizontal orientation, respectively. From these decays the anisotropy as a function of time was determined as

$$r(t) = \frac{I_V(t) - G I_H(t)}{I_V(t) + 2 G I_H(t)},$$

where G is a correction factor compensating for the slightly different instrumental detection sensitivity for vertically and horizontally polarized photons.

Simulations. Resonance energy transfer in the presence of orientational constraints of donor and acceptor dyes was simulated using Mathematica (Wolfram Research) as follows. The orientational dynamics of the fluorophores were modeled as angular diffusion of the dye dipole in a cone[8], approximated by a random walk on the corresponding segment of a unit sphere[9]. For every simulated fluorescence burst, corresponding to an individual rhodanese-GroEL complex, we assumed a random, isotropic, but fixed, relative orientation of the cone axes for donor and acceptor, separated by a distance selected randomly from a normal distribution of given mean and standard deviation. The Brownian dynamics for the diffusive motion of the dye dipoles were combined with a Monte Carlo simulation of the photon emission process of the coupled dye pair[10]. After each time step of 10 ps, the transfer efficiency was calculated from the distance and the relative orientation of donor and acceptor dipoles[11]. For each emitted photon, the orientation of the donor dye at the time of absorption, the orientation of the emitting dye at the time of emission, and the time interval between donor excitation and photon emission were stored for the calculation of anisotropy decays. Each calculation consisted of 5000 simulated bursts of 50 photons each (corresponding to 5000 rhodanese-GroEL complexes with different relative cone orientations) to achieve similar statistics as in the experiment. From the complete data set, transfer efficiency histograms and anisotropy decays were calculated as shown in Figs. 4 and S3.

Supporting Figures

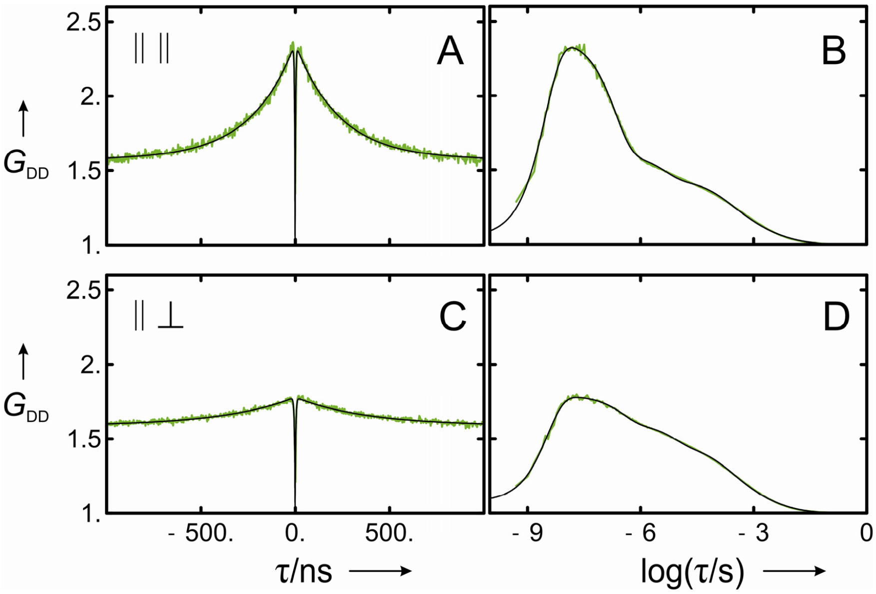


Fig. S1. *Dependence of the donor intensity autocorrelation function of GroEL-bound rhodanese on the polarizations used for signal correlation.* (A,B) Donor (horizontal polarization) – donor (horizontal polarization) intensity correlation of donor-only-labeled rhodanese bound to GroEL. (C,D) Donor (horizontal polarization) – donor (vertical polarization) intensity correlation. The pronounced difference in amplitude indicates that the correlation is caused by rotational motion of the entire rhodanese-chaperone complex[12]. The relative amplitudes are in good agreement with theoretical calculations of the polarization dependence[13] (not shown).

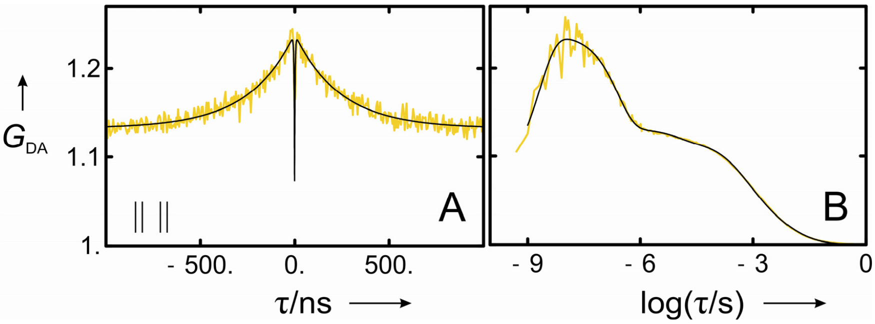


Fig. S2. *Donor (horizontal polarization)-acceptor (horizontal polarization) crosscorrelation of FRET-labeled chaperone-bound rhodanese.* The pronounced correlation on the timescale of rotational diffusion of the entire rhodanese-chaperone complex (A: linear, B: logarithmic scale) shows that the relative orientation of donor and acceptor chromophores in the chaperone-bound protein is rather fixed on the time scale of ~0.2 µs, such that a marked polarization component is retained in spite of energy transfer, implying that the intramolecular distance between the dyes also does not change on this time scale.

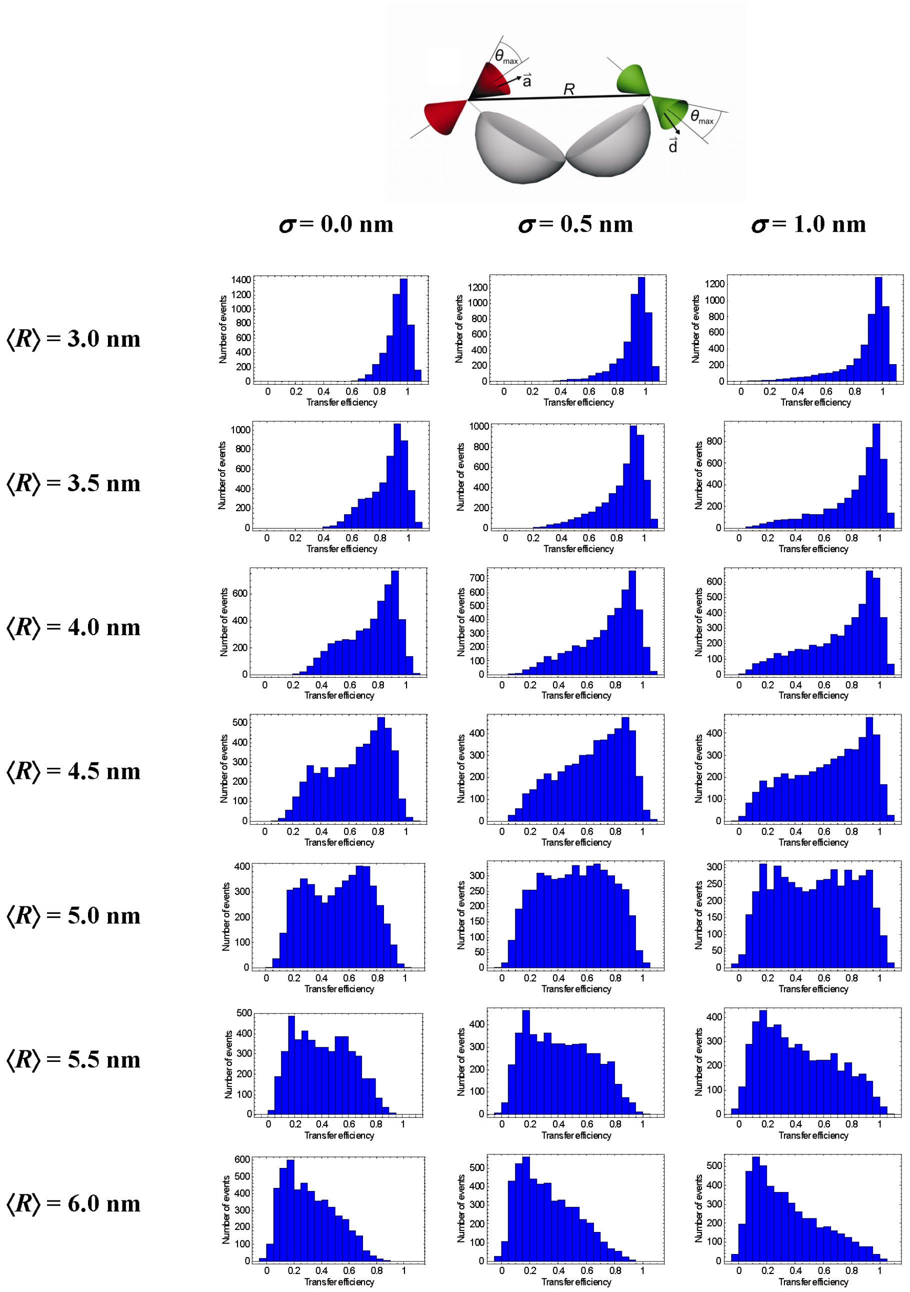


Fig. S3. *Simulated transfer efficiency histograms* with systematic variation of the distance distribution between donor and acceptor, assuming orientational restriction of the dyes as described in the main text and *Experimental Section*. Realistic background values were included for the calculation of the histograms. The shape of the histograms is very sensitive to the mean distance $\langle R \rangle$ and the width (standard deviation σ) of the distance distribution (here assumed to be Gaussian). Note in particular that for very short distances the transfer efficiency distribution is very narrow and close to $E = 1$. The Förster radius R_0 is 5.4 nm.

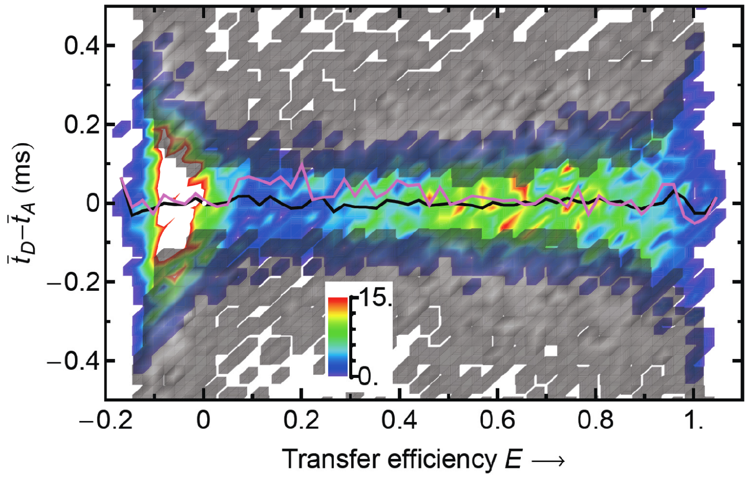


Fig. S4. *Identification of Fluorescence Bursts Affected by Photobleaching.* The burst asymmetry α is shown as a function of the FRET efficiency. The events with $|\alpha_{DA}| > \sigma_{DA}$ are shown in grey, the intensity scale for the remaining events is shown in color. The mean burst asymmetries before and after eliminating the asymmetric events are shown in purple and black, respectively, indicating exclusion of bursts where the acceptor was photobleached during the transit through the confocal volume.

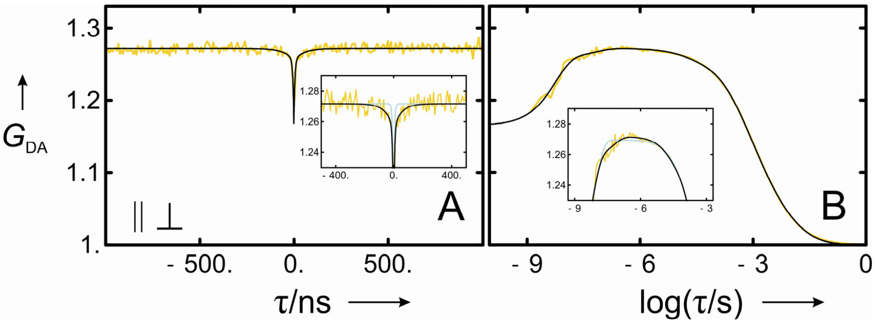


Fig. S5. *Anticorrelation of the Donor (horizontal polarization) - acceptor (vertical polarization) crosscorrelation caused by intramolecular dynamics of rhodanese in 5 M GdmCl.* (A) Short time correlation with linear time base including a fit to the data resulting in a time constant of 80 ns for intramolecular dynamics, in good agreement with the autocorrelation data (Fig. 2 main text). (Inset) Comparison of fits including (black) and excluding (green) distance dynamics. (B) Complete correlation function with logarithmic time base. Fits and inset analogous to (A).

References

[1] F. Hillger, D. Nettels, S. Dorsch, B. Schuler, *J. Fluoresc.* **2007**, *17*, 759-765.

[2] A. L. Horwich, S. G. Burston, H. S. Rye, J. S. Weissman, W. A. Fenton, *Methods Enzymol.* **1998**, *290*, 141-6.

[3] A. M. Bhattacharyya, P. M. Horowitz, *J. Biol. Chem.* **2001**, *276*, 28739-43; J. A. Mendoza, E. Rogers, G. H. Lorimer, P. M. Horowitz, *J. Biol. Chem.* **1991**, *266*, 13044-13049; B. M. Gorovits, P. M. Horowitz, *Biochemistry* **1998**, *37*, 6132-5; G. Zardeneta, P. M. Horowitz, *J Biol Chem* **1992**, *267*, 5811-6.

[4] N. P. Huang, J. Vörös, S. M. De Paul, M. Textor, N. D. Spencer, *Langmuir* **2002**, *18*, 220-230.

[5] B. K. Müller, E. Zaychikov, C. Bräuchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508-3522.

[6] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655-2660.

[7] B. Schuler, *Methods. Mol. Biol.* **2007**, *350*, 115-38.

[8] G. Lipari, A. Szabo, *Biophys. J.* **1980**, *30*, 489-506; K. Kinosita, S. Kawato, A. Ikegami, *Biophys. J.* **1977**, *20*, 289-305; G. F. Schröder, U. Alexiev, H. Grubmüller, *Biophys J* **2005**, *89*, 3757-3770.

[9] J. B. Pedersen, *J. Chem. Phys.* **1972**, *57*, 2680-&.

[10] Z. S. Wang, D. E. Makarov, *J. Phys. Chem. B* **2003**, *107*, 5617-5622.

[11] B. W. Van Der Meer, Coker, G. III, Chen, S. Y. S., *Resonance energy transfer: theory and data*, VCH Publishers, Inc., New York, **1994**.

[12] Ü. Mets, in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer-Verlag, Berlin, **2001**.

[13] P. Kask, P. Piksarv, M. Pooga, Ü. Mets, E. Lippmaa, *Biophys. J.* **1989**, *55*, 213-220.

4 Single-Molecule Spectroscopy of Protein Folding in a Chaperonin Cage

H. Hofmann, F. Hillger, S. H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. A. Lipman, and B. Schuler.

Contributions of Dominik Hänni:

I combined my knowledge about the GroEL/ES system together with my experience in performing single-molecule experiments on surface immobilized molecules to study the folding and the dynamics of rhodanese encapsulated in the GroEL/ES complex. For this I performed for the first time ns-FCS experiments on immobilized molecules. Additionally I tried to directly observe the folding of rhodanese within the GroEL/ES chaperone. Similar as for the bound rhodanese, no observable chain dynamics was found upon encapsulation. An extensive search for folding events in the single-molecule trajectories of the immobilized complexes did not lead to clear results within the accessible timescale.

Single-molecule spectroscopy of protein folding in a chaperonin cage

Hagen Hofmann[a], **Frank Hillger**[a], **Shawn H. Pfeil**[b], **Armin Hoffmann**[a], **Daniel Streich**[a], **Dominik Haenni**[a], **Daniel Nettels**[a], **Everett A. Lipman**[b], **and Benjamin Schuler**[a,1]

[a]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; and [b]Department of Physics, University of California, Santa Barbara, CA 93106

Edited by Arthur L. Horwich, Yale University School of Medicine, New Haven, CT, and approved April 26, 2010 (received for review March 1, 2010)

Molecular chaperones are known to be essential for avoiding protein aggregation in vivo, but it is still unclear how they affect protein folding mechanisms. We use single-molecule Förster resonance energy transfer to follow the folding of a protein inside the GroEL/GroES chaperonin cavity over a time range from milliseconds to hours. Our results show that confinement in the chaperonin decelerates the folding of the C-terminal domain in the substrate protein rhodanese, but leaves the folding rate of the N-terminal domain unaffected. Microfluidic mixing experiments indicate that strong interactions of the substrate with the cavity walls impede the folding process, but the folding hierarchy is preserved. Our results imply that no universal chaperonin mechanism exists. Rather, a competition between intra- and intermolecular interactions determines the folding rates and mechanisms of a substrate inside the GroEL/GroES cage.

chaperone | confinement | microfluidic mixing | FRET | fluorescence

In the recent past, a large number of components have been identified that control and modulate protein folding in vivo. This machinery includes molecular chaperones (1–3), sophisticated quality control systems, and complex mechanisms for protein translocation and degradation (3, 4), reflecting the importance of regulating the delicate balance of protein folding, misfolding, and aggregation in the cell. Such cellular factors exert conformational constraints on protein molecules that are expected to have a strong effect on the corresponding free-energy surfaces for folding (5). However, while the combination of cellular, biochemical, and structural data has led to some plausible qualitative models for the processes involved, mechanistic investigations comparable to those of autonomous protein folding in vitro (5–8) have been complicated by the complexity of the systems and the conformational heterogeneity involved (9). Even the autonomous folding of chaperone substrate proteins has been difficult to investigate because of their strong aggregation tendency (10). Contributions from confinement and crowding have been addressed in numerous studies using molecular simulations and theory (11–20), but many of these concepts have eluded experimental examination.

Here, we take a step towards closing this gap by investigating the GroEL/GroES chaperonin (1–3, 9) with single-molecule fluorescence spectroscopy (21–24), a method that is starting to provide previously inaccessible information on chaperone-mediated protein folding (25–30). GroEL/GroES is a remarkable molecular machine that binds nonnative proteins and allows them to fold within a cavity formed by the heptameric rings of GroEL and GroES. However, the cavity is only slightly larger than the folded structure of typical proteins known to interact with the chaperonin. The large volume of unconfined unfolded protein chains compared to the size of the cavity raises the question of whether and how such strong confinement affects the folding reaction (12–16, 18, 31, 32). By labeling the classic substrate protein rhodanese (33) with donor and acceptor fluorophores, we can follow the folding reaction with multiparameter single-molecule Förster resonance energy transfer (FRET) (34) and probe the folding pathway of rhodanese inside and outside the chaperonin cavity in much greater detail than previously possible.

Results and Discussion

Chaperone-Mediated Protein Folding Observed with Single-Molecule FRET. To achieve an optimal discrimination of native and nonnative conformations, three variants of the two-domain protein rhodanese were investigated. Two fluorophores (Alexa Fluor 488 and Alexa Fluor 594) were attached to each variant to map the folding of the N-terminal domain (N variant), the structure formation of the linker separating both domains (L variant), and the folding of the C-terminal domain (C variant) (Fig. 1*A*). Fig. 1*B* shows histograms of the transfer efficiency E for all three rhodanese variants, determined from photon bursts of individual rhodanese molecules freely diffusing through the observation volume of the confocal instrument. Under native conditions (Fig. 1*B*), two peaks are observed for each variant: the peaks at $E = 0.67$ for the N variant, $E = 0.69$ for the L variant (27), and $E = 0.98$ for the C variant result from native rhodanese molecules; the peaks near $E = 0$ result from molecules lacking an active acceptor dye and can be eliminated by dual color excitation of donor and acceptor (35, 36) (Fig. 1*B*, see *SI Appendix* for details). When the refolding of rhodanese is initiated in the presence of GroEL, rhodanese binds to the chaperonin ring, resulting in characteristic broad transfer efficiency distributions for all three variants with maxima at $E > 0.8$ (Fig. 1*B*), whose width originates predominantly from static orientational heterogeneity of the fluorophores (27). To investigate refolding inside the chaperonin cavity, we use the single ring variant of GroEL (SR1), which resembles the folding active state of GroEL, but does not release the substrate protein (37, 38). Upon incubation of the SR1-rhodanese complex with ATP and the cochaperone GroES, stable complexes assemble (*SI Appendix: Fig. S1*), and rhodanese is displaced into the cavity formed by SR1 and GroES within a few seconds (31, 37). During the folding of rhodanese inside the chaperonin cage, we observe that the transfer efficiency histograms of all variants approach the histograms of the free native state (Fig. 1*B*). The concurrent decrease in the donor and acceptor fluorescence anisotropies (*SI Appendix: Fig. S2*) indicates an increase in rotational freedom of the fluorophores during folding, a behavior observed previously during release of substrate proteins into the cavity (31, 37). Even though some residual broadening from static orientational heterogeneity of the fluorophores remains in the encapsulated folded state, the characteristic changes in the transfer efficiency histograms allow us to follow the folding of rhodanese inside the

---

Author contributions: H.H. and B.S. designed research; H.H. and F.H. performed research; F.H., S.H.P., A.H., D.S., D.H., D.N., and E.A.L. contributed new reagents/analytic tools; H.H. analyzed data; and H.H. and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. Email:schuler@bioc.uzh.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002356107/-/DCSupplemental.

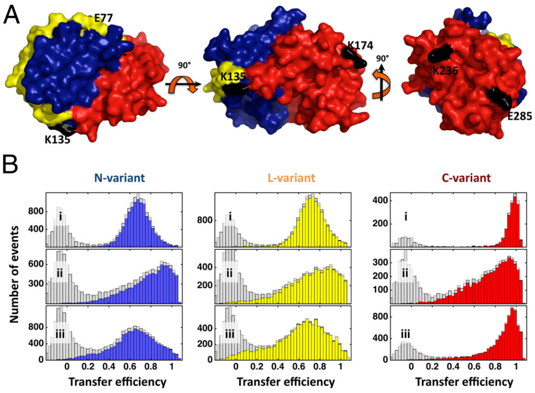


Fig. 1. Native structure and transfer efficiency histograms of the rhodanese variants. (*A*) Surface representation of rhodanese showing the N-terminal domain (*blue*), the interdomain linker (*yellow*), and the C-terminal domain (*red*) (protein data bank entry 1rhs). The rhodanese variants E77C/K135C (*N variant*), K135C/K174C (*L variant*), and K236C/E285C (*C variant*) were labeled with Alexa Fluor 488 as a donor and Alexa Fluor 594 as an acceptor. Label attachment sites are indicated in *black*. (*B*) Transfer efficiency histograms of native rhodanese (*i*), the SR1-rhodanese complex (*ii*), and the SR1-rhodanese complex 1.5 h after addition of GroES and ATP (*iii*). The gray histograms were recorded with donor excitation only; the colored histograms were recorded using dual color excitation of donor and acceptor (35, 73) to eliminate the contribution close to $E = 0$ from molecules lacking an active acceptor dye.

GroEL cage and compare the kinetics with those of its autonomous refolding in solution.

Folding Kinetics Outside and Inside the Cage. Autonomous refolding of rhodanese was initiated by manually diluting unfolded rhodanese (4 M guanidinium chloride) into native conditions. Chaperone-mediated refolding was triggered by mixing the binary rhodanese-SR1 complex with GroES (1 μM) and ATP (2 mM). The single-molecule fluorescence signal was recorded until the refolding reaction was complete, typically for 2 h. To obtain time-resolved FRET efficiency histograms, we performed a moving window analysis by splitting the single-molecule recordings into short intervals of equal duration (50 s for the autonomous folding of the C variant and 300 s for all other variants) (Fig. 2*A*, see *SI Appendix* for details). For both the autonomous and the SR1-mediated folding reactions of all variants, transfer efficiency histograms characteristic of native molecules developed during the measurements (Fig. 2*A*).

For the autonomous folding of rhodanese, it might appear feasible to analyze the kinetics of the histograms in terms of two well-defined states, but the broad histograms of the chaperone-mediated reaction obviously require a less model-dependent analysis. We thus employed singular value decomposition (SVD), which can be used to factorize a matrix representation of the experimental data into a minimal set of basis vectors and amplitude vectors, whose linear combination, weighted by the corresponding singular values, can be used to represent the data (39). In our case, we can analyze the change of the two-dimensional histograms with time to determine the kinetics and the minimum number of distinguishable molecular species required for describing the folding process without loss of information (see *SI Appendix* for details). All nine experimentally accessible observables, represented in two-dimensional histograms (*SI Appendix: Fig. S3–5*), were combined in one global SVD analysis (Fig. 3*C*, *D* and *SI Appendix: Fig. S6–8*). While parameters such as transfer efficiency and burst duration are more sensitive to global changes in the dimension of the protein, fluorescence lifetime and fluorescence anisotropy report on changes in the local environment and the rotational freedom of the fluorophores, respectively. The basis vectors contain information about the parts of the histograms that change over time, and the amplitude vectors report on the corresponding kinetics. Fig. 3 *C* and *D* show examples of the multidimensional SVD for the autonomous and chaperone-mediated folding reaction of the L variant at 24 °C. Interestingly, for all observables, the signal change is dominated by the first two SVD components (Fig. 3 *C* and *D* and *SI Appendix: Fig. S6–9*). The first component captures mainly an increase in the molecular brightness over time (Fig. 3 *A*, *B*, S6–8), which is probably caused by the burial of tryptophan residues in the native structure that quench donor and acceptor in the denatured state (40). The second component corresponds to the changes in all other spectroscopic parameters, e.g., the transfer efficiency (Fig. 3 and *SI Appendix: Fig. S6–8*). The two SVD components yielded very similar rate constants for each of the individual folding reactions (Fig. 2*B*), indicating the dominance of two distinguishable molecular species*. In all cases, the SVD amplitude vectors were well described by single exponential relaxations.

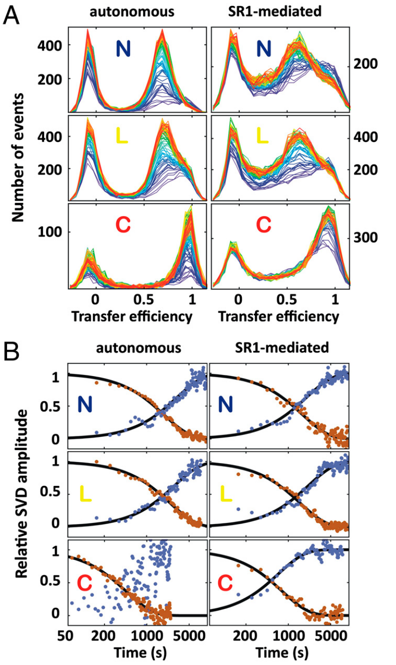


Fig. 2. Kinetic analysis of the autonomous and chaperone-mediated rhodanese refolding reactions using SVD. (*A*) Transfer efficiency histograms as a function of time (*progressing from blue to red*) of the autonomous (*Left*) and SR1-mediated (*Right*) folding reaction for N variant, L variant, and C variant (from *Top* to *Bottom*) at 24 °C. (*B*) Kinetics from the first (*red*) and second (*blue*) amplitude vectors of the SVD for the autonomous (*Left*) and SR1-mediated (*Right*) folding reaction of the N variant, L variant, and C variant.

A comparison of the resulting rate constants for the autonomous folding of the rhodanese variants (Fig. 2*B*) suggests a simple folding mechanism. The folding rate constant of the N-terminal domain $(4.2 \pm 1.4)10^{-4}\ \mathrm{s}^{-1}$ coincides with the formation of the native linker structure $(3.9 \pm 1.2)10^{-4}\ \mathrm{s}^{-1}$, but the C-terminal domain folds almost six times faster $(2.3 \pm 0.6)10^{-3}\ \mathrm{s}^{-1}$ at 24 °C, indicating that the C-terminal domain folds prior to the N-terminal

---

*For the autonomous folding of the C variant, only the first component contains kinetic information because of the overlap of native and nonnative populations in the histograms.

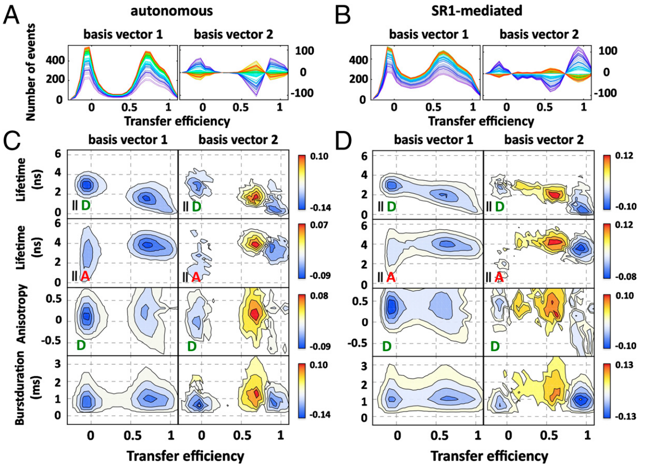


Fig. 3. Examples of basis vectors from multidimensional SVD for the autonomous and SR1-mediated folding reactions of the L variant. (*A*, *B*) Time evolution (*progressing from blue to red*) of the first (*Left*) and second (*Right*) one-dimensional-basis vectors for the autonomous (*A*) and SR1-mediated folding (*B*) of the L variant. Note that the one-dimensional-basis vectors shown here are just one possible projection of the multidimensional basis vectors on the transfer efficiency dimension to illustrate the kinetics. (*C*, *D*) Examples of two-dimensional-basis vectors from multidimensional SVD for the autonomous (*C*) and SR1-mediated (*D*) folding reactions of the L variant (*from Top to Bottom*: donor and acceptor fluorescence lifetime, donor fluorescence anisotropy, duration of bursts). The color code indicates the absolute SVD amplitude (*see color scale*). The basis vectors indicate the positions of changes of the corresponding observables in the histograms and are ordered according to their singular values.

domain. This interpretation is further corroborated by limited proteolysis experiments (*SI Appendix*: *Fig. S10*), excluding a dominant effect of the fluorophores on the folding mechanism[†]. Interestingly, the folding hierarchy of the domains is preserved in the highly confined space of the GroEL/GroES complex. Correspondingly, the basis vectors for autonomous and chaperone-mediated folding are similar (Fig. 3 and *SI Appendix*: *Fig. S6–8*). However, the folding rates of the domains are affected differently by the chaperonin cavity: the folding rate constant of the N-terminal domain $(4.5 \pm 1.2)10^{-4}\ \mathrm{s}^{-1}$ and the rate constant for formation of the native linker $(5.5 \pm 1.1)10^{-4}\ \mathrm{s}^{-1}$ are not changed by the chaperonin environment (Fig. 2*B*, 4*A*). In contrast, folding of the C-terminal domain is decelerated by a factor of two inside the chaperonin cavity at 24 °C $(1.0 \pm 0.4)10^{-3}\ \mathrm{s}^{-1}$ (Fig. 2*B*, 4*A*); this effect increases to an eightfold deceleration when extrapolated to 37 °C (Fig. 4*A*). Even though our values for the rate constants (*SI Appendix*: *Fig. S11*) lie within the range of previous results obtained by enzymatic assays, a rigorous comparison to published results is complicated by the considerable spread of the rate constants reported (10, 32, 37, 41, 42). Possible reasons for this variability are the pronounced sensitivity of the system to solution conditions (10, 37), temperature (Fig. 4*A*), and difficulties in completely eliminating aggregation at the protein concentrations required in ensemble experiments (10). Importantly, aggregation in single-molecule experiments is not only improbable because of the low protein concentrations used, but it can also be monitored stringently in situ (43) and can thus be excluded for all measurements (*SI Appendix*: *Fig. S12*).

Rapid Kinetics from Microfluidic Mixing. A complete picture of the conformational dynamics of rhodanese during refolding requires an investigation on all biologically relevant time scales from milliseconds to hours. While the time range of minutes to hours is accessible with the above-described manual mixing experiments, recent advances in the development of continuous-flow microfluidic mixing devices (44–47) allow us to study folding reactions on the single-molecule level from milliseconds to seconds. Here, we use a microfluidic mixer designed specifically for single-molecule measurements of fast protein folding kinetics (46). A sample solution in the inlet channel containing SR1-bound rhodanese (Ch2, Fig. 5*A*) was mixed with buffer containing ATP and GroES that entered from the two side channels (Ch1 and Ch3, Fig. 5*A*). By placing the confocal volume at different positions along the observation channel (Ch4), we obtained transfer efficiency histograms at different times after initiation of the reaction (see *SI Appendix*). Mixing the SR1-rhodanese complexes with 2 mM ATP and 2 μM GroES results in complete binding of GroES to SR1 in ~200 ms (48), which triggers the release of the substrate protein into the chaperonin cavity. Active unfolding of the substrate protein driven by the conformational changes of the apical domains of GroEL upon binding of ATP and GroES has been proposed to support protein folding (28, 49, 50). Surprisingly, we observed no obvious changes in the transfer efficiency histograms on a timescale from milliseconds to seconds (Fig. 5*B*). Only by averaging over the entire transfer efficiency histograms, we obtained a slight change in transfer efficiency of both variants by 0.05 ± 0.01 (Fig. 5*C*). The rate constant for the initial decrease of $7 \pm 2\ \mathrm{s}^{-1}$ is close to the value reported for GroES-binding ($19\ \mathrm{s}^{-1}$) under these conditions (48), and the slower increase can be described with the reported rate of apical domain movement of SR1 under substrate load of $0.68\ \mathrm{s}^{-1}$ (48). These changes in the average transfer efficiency probably reflect very small conformational rearrangements of the substrate or the fluorophores during encapsulation, which are unlikely to be able to cause a selective deceleration of folding of the C-terminal domain inside the chaperonin cavity on the time scale of minutes to hours. We thus need to investigate alternatives for the molecular basis of the effect of the chaperonin on rhodanese folding.

Possible Contributions to the Folding Rates in the Chaperonin Cage. Changes in folding rate constants can be caused by several effects. As a starting point, we express the folding rate constant k in terms of a generalized reaction rate equation,

$$k = k_0 \exp(-\Delta G^{\ddagger}/RT) = k_0 \exp(-\Delta H^{\ddagger}/RT + \Delta S^{\ddagger}/R), \qquad [1]$$

where $\Delta G^{\ddagger}$ is the height of the free-energy barrier separating the denatured from the native state, R is the gas constant, and T is the absolute temperature. The preexponential factor k_0 sets a "speed limit" (7) for the reaction and can be thought of as an attempt frequency for crossing the barrier (6).

First, we address the possibility that the decelerated folding of the C-terminal domain in the chaperonin is caused by an increase in barrier height. Since the free-energy barrier is not accessible directly, we investigate the enthalpic and entropic contributions to $\Delta G^{\ddagger}$ separately. The activation enthalpy $\Delta H^{\ddagger}$ can be obtained from the temperature dependence of the folding rate constants. Surprisingly, the rate-limiting step of rhodanese folding, i.e., folding of the N-terminal domain and formation of the native interdomain linker

---

[†] For further discussion of the effects of fluorophore labeling on the folding reaction, see *SI Appendix: Materials and Methods and Figs. S2 and S11* .

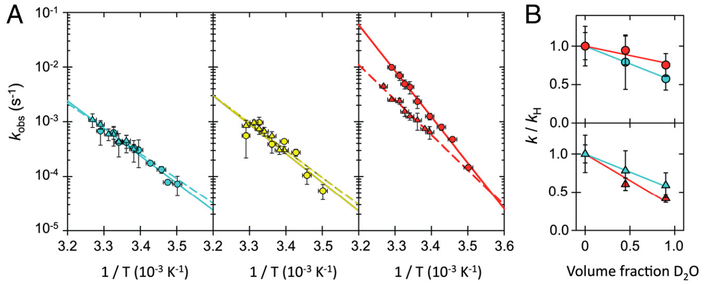


Fig. 4. Effect of temperature and solvent entropy on the autonomous and SR1-mediated folding reactions. (*A*) Arrhenius plots for the autonomous (*Circles*) and SR1-mediated (*Triangles*) folding reaction for the N variant (*Left*), L variant (Center), and C variant (*Right*). *Solid* (autonomous) and *dashed* (SR1-mediated) *lines* are Arrhenius fits according to Eq. **1**. Error bars indicate standard deviations estimated from the two SVD-components or from two or three independent measurements for the cases where several measurements were available. The resulting activation enthalpies $\Delta H^{\ddagger}$ are (96 ± 7) kJ mol^{-1} (autonomous) and (88 ± 8) kJ mol^{-1} (SR1-mediated) for the N variant, (100 ± 25) kJ mol^{-1} (autonomous) and (100 ± 17) kJ mol^{-1} (SR1-mediated) for the L variant, and (161 ± 5) kJ mol^{-1} (autonomous) and (123 ± 7) kJ mol^{-1} (SR1-mediated) for the C variant. (*B*) Kinetic solvent isotope effects shown by the dependence of the ratio k/k_H on the volume fraction of D_2O at 27 °C for the autonomous (*Top*) and SR1-mediated (*Bottom*) refolding rates of the N variant (*cyan*) and C variant (*red*). Error bars indicate standard deviations estimated from at least two independent measurements, and lines represent linear regressions to illustrate the trends.

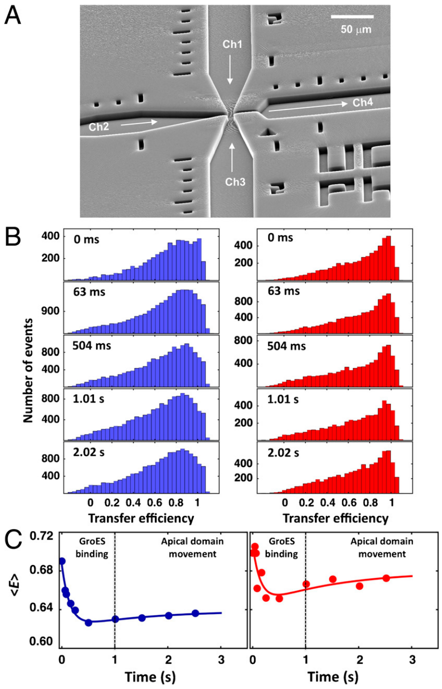


Fig. 5. Rapid processes in SR1-mediated rhodanese folding investigated with microfluidic mixing. (*A*) Scanning electron micrograph of the microfluidic mixing device (46). SR1-bound rhodanese in Ch2 is mixed with GroES and ATP in Ch1 and Ch3 in the narrow mixing region. Measurements were taken at different positions along the observation channel Ch4, corresponding to different times after mixing. (*B*) Transfer efficiency histograms of SR1-bound N variant (*Left*) and C variant (*Right*) at different times after mixing GroEL-bound rhodanese with 2 μM GroES and 2 mM ATP. (*C*) Kinetics of the average transfer efficiency $\langle E \rangle$ for the SR1-bound N variant (*Left*) and C variant (*Right*) obtained from the histograms in *B*. The lines represent a global double exponential fit to the data. The rate constant describing the slow increase after the initial drop was constrained to the rate constant of the apical domain movement of 0.68 s^{-1} (48). The histograms were recorded using dual excitation of donor and acceptor (35, 73) to eliminate the contribution close to $E = 0$ from molecules lacking an active acceptor dye.

conformation are not affected by the chaperonin over the entire accessible temperature range (Fig. 4*A*). Assuming Arrhenius behavior, we find the activation enthalpies of the chaperonin-mediated folding reactions to be indistinguishable within experimental error from those of the autonomous reaction. However, folding of the C-terminal domain is slower in the chaperonin than free in solution at all temperatures [‡] (51), with a significantly lower activation enthalpy $(123 \pm 7\ \mathrm{kJ\,mol^{-1}})$ compared to the autonomous reaction $(161 \pm 5\ \mathrm{kJ\,mol^{-1}})$ (Fig. 4*A*). An increase in the enthalpic contribution to the free-energy barrier can thus be excluded as a cause of the slower folding of the C-terminal domain in the cavity.

The second possible origin of a change in $\Delta G^{\ddagger}$ is a change in activation entropy, $\Delta S^{\ddagger}$, upon encapsulation. The most important contributions to $\Delta S^{\ddagger}$ are conformational entropy and the entropy of solvation. Confinement in the chaperonin is expected to reduce the conformational entropy of the denatured state (12–16, 18). Consequently, the difference in conformational entropy of the denatured state and the transition state should decrease in the chaperonin cavity, which would reduce the height of the free-energy barrier and thus accelerate folding (12–16, 18), the opposite of what we observe. Conformational entropy is thus unlikely to be the cause of slower folding inside the chaperone.

Recent theoretical work suggests an important role of confined water molecules in chaperonin-assisted protein folding (52). We investigated this possibility by means of kinetic solvent isotope effects caused by replacing H_2O by D_2O in the samples. The stronger hydrogen bonding in D_2O is thought to increase the hydrophobic effect and stabilize proteins (53–57). If water dominated the entropy change during the chaperone-mediated folding reaction, the kinetic solvent isotope effect in the chaperone should be significantly different from that of the autonomous folding reaction. Fig. 4*B* shows the dependence of the ratio k/k_H on the volume fraction of D_2O in the buffer, where k is the refolding rate constant at different fractions of D_2O, and k_H is the rate constant in water. The rate constants for autonomous folding of both the N- and C variants of rhodanese were reduced by a factor of 1.5 to 2 in 90% D_2O. A similar decrease in the folding rate constants was found for the chaperonin-mediated folding reactions of both variants, making the presence of confined water molecules an unlikely cause of a change in folding rates in the GroEL/GroES cavity.

[‡] At temperatures above 35 °C, the GroEL/GroES-rhodanese complex tends to aggregate even at single-molecule concentrations, while at temperatures lower than 18 °C, the GroEL oligomer dissociates (51). For spontaneous rhodanese folding, the temperature range is limited by the freezing point of water at low temperature and increased quenching of the dye molecules at higher temperature, leading to poor data quality above 35 °C.

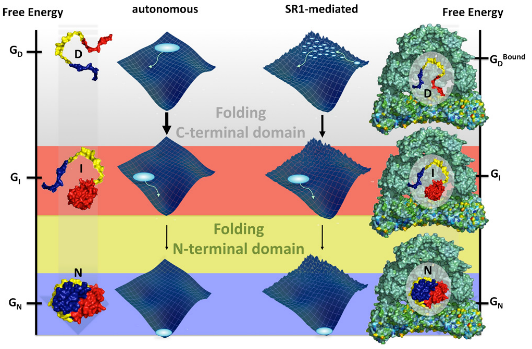


Fig. 6. Schematic of the autonomous and SR1-mediated folding of rhodanese. Rhodanese folds via a partially folded intermediate, in which the C-terminal domain (*red*) is already folded (*Left*). In the chaperonin-mediated reaction, molecular friction (energetic roughness) from interactions with the cavity wall decelerates the folding of the C-terminal domain. However, the folding pathway of rhodanese is preserved (*Right*).

In summary, we find no indication that an increase in the free-energy barrier height is the origin of the slower folding of the C-terminal domain we observed in the chaperonin cage. Alternatively, our observation may originate from effects that essentially enter into the preexponential factor k_0 in Eq. **1**. In Kramers-type theories of protein folding (6, 58, 59), k_0 is expressed in terms of an effective intramolecular diffusion coefficient D of the polypeptide with $k_0 \propto D$, where D can be related to the roughness of the underlying free energy surface for folding (60). The reason for the lower folding rate in this picture is a decrease in the effective mobility of the polypeptide chain, which reduces the rate at which new configurations can be explored (61–63). The origin of such molecular friction can be both intra- and intermolecular interactions. For chaperonin-mediated folding, this would correspond to nonnative interactions within the folding polypeptide and interactions with the walls of the cavity.

A considerable body of theoretical work suggests that, even though moderate confinement of a polypeptide in a cavity can accelerate folding entropically, reduced folding rates are expected from stronger confinement that restricts conformational fluctuations and leads to an increase in molecular friction (13–19). In view of the small size of the chaperonin cavity, resembling a sphere with a radius of ~3 nm (assuming a cavity volume of 120 nm^3 (64)), compared to the radius of gyration of denatured rhodanese of ~3.8 nm (calculated assuming the typical persistence length of 0.4 nm for unfolded protein chains under native conditions (65)), a significant effect of confinement on the folding dynamics of rhodanese may be expected. However, confinement alone should influence the folding rates of both domains to a similar extent, in contrast to our experimental observation, implying an additional influence of interactions of the substrate with the cavity walls (20, 31). Recent theoretical work indicates a pronounced effect of the interaction strength between cage and protein on folding rates: moderate interaction strengths can, in a narrow range, accelerate folding by iterative binding and dissociation events, but simulations predict a deceleration of folding for strong interactions (11, 16, 66). Evidence for such interactions comes from our microfluidic mixing experiments (Fig. 5*B*), which indicate a lack of conformational rearrangements of rhodanese during encapsulation, and thus suggest that interactions between rhodanese and GroEL persist in the encapsulated state. Even the first histograms from the manual mixing experiments (Fig. 2*A*) still resemble those of the binary SR1-rhodanese complex (Fig. 1*B*). Recent cryoelectron microscopy experiments show that substrate proteins bound to GroEL are predominantly localized deep inside the cavity (67, 68), a situation that will facilitate interactions with the chaperonin walls in the GroEL-bound state. The particularly strong interactions of rhodanese with GroEL (69, 70) are thus likely to increase molecular friction of the substrate protein in the cavity. If we assume that the effect of such interactions can be approximated by an effective dissociation step from the chaperonin wall, protein-chaperone interactions will become rate-limiting for faster processes, such as folding of the C-terminal domain, whereas slower processes, such as folding of the N-terminal domain, will be much less affected, in agreement with our observations (Fig. 6). A further understanding of the molecular basis of these effects will benefit greatly from the increasingly detailed information available from theory and simulations (20).

Our results illustrate how multiparameter single-molecule spectroscopy in combination with microfluidic mixing opens a new opportunity for identifying previously elusive effects of molecular chaperones on protein folding mechanisms. Major advantages of the approach are the availability of distributions of observables instead of mean values, the complementarity of the different types of spectroscopic information from a single measurement that can be used for a global analysis of all observables, the broad range of time scales accessible, and the extremely low protein concentrations employed, which allow aggregation to be excluded from affecting the folding kinetics. Although the biological function of the GroEL/GroES system is suggestive of an acceleration of folding rates, our results show that chaperonins can even slow down protein folding processes, and support the view that preventing aggregation of proteins is more important for cellular viability than accelerating protein folding reactions (71). However, our observations call for a differentiated view of chaperone action: since the folding rates of the domains within a single protein can be affected differently by the chaperonin, it is improbable that there is one universal chaperonin mechanism at work. This notion is supported by the large variability of effects of chaperonins on the folding of different proteins reported in the literature (20) and by theoretical concepts that provide a quantitative framework for the competition between intra- and intermolecular interactions that determine the folding rate and mechanism of a substrate protein inside the GroEL/GroES cage (1, 8, 11–20, 66). Future experimental and theoretical investigations will have to address the potential synergies of the different mechanisms, whose subtle balance may be required to achieve the promiscuity of many molecular chaperones.

Materials and Methods

Expression and purification of SR1 (72) and preparation and labeling of cysteine variants of rhodanese (27) were performed as described previously. Binary complexes of SR1 and rhodanese were prepared by diluting unfolded rhodanese (in 4 M guanidinium chloride) into 50 mM TrisHCl, 10 mM $MgCl_2$, 5 mM KCl, 100 mM 2-mercaptoethanol, 0.001% Tween 20 (Pierce), pH 7.5 (folding buffer) containing 1 μM SR1. The complex was purified using size exclusion chromatography. Single-molecule fluorescence experiments were performed with a MicroTime 200 confocal microscope (PicoQuant). The temperature was adjusted with a Peltier-controlled sample holder. Autonomous and SR1-mediated refolding of rhodanese were performed in folding buffer. Data reduction for the refolding kinetics was performed by global analysis of all observables using SVD. For rapid mixing experiments, microfluidic mixers fabricated by replica molding in polydimethylsiloxane were used (46). For detecting the GroES-ATP-mediated encapsulation reaction of the SR1-bound rhodanese variants in the microfluidic mixer, the binary rhodanese-SR1 complex was mixed at a volume ratio of 1 : 5.7 with 2.4 μM GroES and 2.4 mM ATP, resulting in final concentrations of 2 μM GroES and 2 mM ATP. See *SI Appendix* for details.

ACKNOWLEDGMENTS. We thank J. Enderlein for valuable suggestions, G. Lorimer for the gift of the SR1-plasmid, and K. Marquardt (Center for Microscopy and Image Analysis, University of Zurich) for the scanning electron micrographs. We thank P. Schütz for simulations and discussion. This work was supported by a Starting Independent Researcher Grant of the European Research Council (to B.S.), the Swiss National Center for Competence in Research for Structural Biology (to B.S.), the Swiss National Science Foundation (to B.S.), the VolkswagenStiftung (to B.S.), the Human Frontier Science

Program (to E.A.L., B.S.), the Defense Microelectronics Activity (DMEA) Center for Nanoscience Innovation for Defense (to E.A.L.), and the Forschungskredit of the University of Zurich (to B.S.). A portion of this work was done in the University of California, Santa Barbara (UCSB) nanofabrication facility, part of the National Science Foundation-funded National Nanotechnology Infrastructure Network. E.A.L. is an Alfred P. Sloan research fellow.

1. Thirumalai D, Lorimer GH (2001) Chaperonin-mediated protein folding. *Annu Rev Biophys Biomol Struct* 30:245–269.
2. Hartl FU, Hayer-Hartl M (2002) Protein folding—Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295:1852–1858.
3. Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. *Cell* 125:443–451.
4. Ellgaard L, Helenius A (2003) Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* 4:181–191.
5. Fersht AR (1998) *Structure and mechanism in protein science* (W.H. Freeman and Company, New York).
6. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21:167–195.
7. Eaton WA, et al. (2000) Fast kinetics and mechanisms in protein folding. *Annu Rev Biophys Biomol Struct* 29:327–359.
8. Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys* 39:159–183.
9. Fenton WA, Horwich AL (2003) Chaperonin-mediated protein folding: fate of substrate polypeptide. *Quarterly Reviews of Biophysics* 36:229–256.
10. Apetri AC, Horwich AL (2008) Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proc Natl Acad Sci USA* 105:17351–17355.
11. Betancourt MR, Thirumalai D (1999) Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J Mol Biol* 287:627–644.
12. Zhou HX, Dill KA (2001) Stabilization of proteins in confined spaces. *Biochemistry* 40:11289–11293.
13. Klimov DK, Newfield D, Thirumalai D (2002) Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proc Natl Acad Sci USA* 99:8019–8024.
14. Baumketner A, Jewett A, Shea JE (2003) Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. *J Mol Biol* 332:701–713.
15. Takagi F, Koga N, Takada S (2003) How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc Natl Acad Sci USA* 100:11367–11372.
16. Jewett AI, Baumketner A, Shea JE (2004) Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. *Proc Natl Acad Sci USA* 101:13192–13197.
17. Cheung MS, Klimov D, Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA* 102:4753–4758.
18. Mittal J, Best RB (2008) Thermodynamics and kinetics of protein folding under confinement. *Proc Natl Acad Sci USA* 105:20233–20238.
19. Zhou HX (2008) Protein folding in confined and crowded environments. *Arch Biochem Biophys* 469:76–82.
20. Jewett A, Shea J (2009) Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell Mol Life Sci* 67:255–276.
21. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
22. Michalet X, Weiss S, Jäger M (2006) Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* 106:1785–1813.
23. Borgia A, Williams P, Clarke J (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* 77:101–125.
24. Haran G (2003) Single-molecule fluorescence spectroscopy of biomolecular folding. *J Phys Cond Mat* 15:R1291–R1317.
25. Yamasaki R, et al. (1999) Single molecular observation of the interaction of GroEL with substrate proteins. *J Mol Biol* 292:965–972.
26. Taguchi H, Ueno T, Tadakuma H, Yoshida M, Funatsu T (2001) Single-molecule observation of protein-protein interactions in the chaperonin system. *Nat Biotechnol* 19:861–865.
27. Hillger F, et al. (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angewandte Chemie International Edition* 47:6184–6188.
28. Sharma S, et al. (2008) Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* 133:142–153.
29. Frank GA, et al. (2010) Out-of-equilibrium conformational cycling of GroEL under saturating ATP concentrations. *Proc Natl Acad Sci USA* 107:6270–6274.
30. Mickler M, Hessling M, Ratzke C, Buchner J, Hugel T (2009) The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. *Nat Struct Mol Biol* 16:281–286.
31. Tang YC, et al. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* 125:903–914.
32. Farr GW, Fenton WA, Horwich AL (2007) Perturbed ATPase activity and not "close confinement" of substrate in the *cis* cavity affects rates of folding by tail-multiplied GroEL. *Proc Natl Acad Sci USA* 104:5342–5347.
33. Mendoza JA, Rogers E, Lorimer GH, Horowitz PM (1991) Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. *J Biol Chem* 266:13044–13049.
34. Widengren J, et al. (2006) Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal Chem* 78:2039–2050.
35. Kapanidis AN, et al. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci USA* 101:8936–8941.
36. Müller BK, Zaychikov E, Bräuchle C, Lamb DC (2005) Pulsed interleaved excitation. *Biophys J* 89:3508–3522.
37. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL (1996) Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* 84:481–490.
38. Weissman JS, et al. (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* 83:577–587.
39. Henry E, Hofrichter J (1992) Singular value decomposition—Application to analysis of experimental data. *Methods Enzymol* 210:129–192.
40. Doose S, Neuweiler H, Sauer M (2009) Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *Chemphyschem* 10:1389–1398.
41. Brinker A, et al. (2001) Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* 107:223–233.
42. Chaudhry C, et al. (2003) Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *EMBO J* 22:4877–4887.
43. Hillger F, Nettels D, Dorsch S, Schuler B (2007) Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc* 17:759–765.
44. Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.
45. Nguyen NT, Wu ZG (2005) Micromixers—a review. *J Micromech Microeng* 15:R1–R16.
46. Pfeil SH, Wickersham CE, Hoffmann A, Lipman EA (2009) A microfluidic mixing system for single-molecule measurements. *Rev Sci Instrum* 80:055105.
47. Lemke EA, et al. (2009) Microfluidic device for single-molecule experiments with enhanced photostability. *J Am Chem Soc* 131:13610–13612.
48. Motojima F, Chaudhry C, Fenton WA, Farr GW, Horwich AL (2004) Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc Natl Acad Sci USA* 101:15005–15012.
49. Lin Z, Madan D, Rye HS (2008) GroEL stimulates protein folding through forced unfolding. *Nat Struct Mol Biol* 15:303–311.
50. Shtilerman M, Lorimer GH, Englander SW (1999) Chaperonin function: folding by forced unfolding. *Science* 284:822–825.
51. Panda M, Horowitz PM (2004) Activation parameters for the spontaneous and pressure-induced phases of the dissociation of single-ring GroEL (SR1) chaperonin. *Protein J* 23:85–94.
52. England J, Lucent D, Pande V (2008) Rattling the cage: computational models of chaperonin-mediated protein folding. *Curr Opin Struct Biol* 18:163–169.
53. Kresheck GC, Schneider H, Scheraga HA (1965) The effect of D_2O on the thermal stability of proteins. Thermodynamic parameters for the transfer of model compounds from H_2O to D_2O. *J Phys Chem* 69:3132–3144.
54. Makhatadze GI, Clore GM, Gronenborn AM (1995) Solvent isotope effect and protein stability. *Nat Struct Biol* 2:852–855.
55. Parker MJ, Clarke AR (1997) Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry* 36:5786–5794.
56. Lopez MM, Makhatadze GI (1998) Solvent isotope effect on thermodynamics of hydration. *Biophys Chem* 74:117–125.
57. Dougan L, Koti AS, Genchev G, Lu H, Fernandez JM (2008) A single-molecule perspective on the role of solvent hydrogen bonds in protein folding and chemical reactions. *Chemphyschem* 9:2836–2847.
58. Kramers H (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7:284–304.
59. Socci ND, Onuchic JN, Wolynes PG (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 104:5860–5868.
60. Zwanzig R (1988) Diffusion in a rough potential. *Proc Natl Acad Sci USA* 85:2029–2030.
61. Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103:12394–12399.
62. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
63. Cellmer T, Henry E, Hofrichter J, Eaton W (2008) Measuring internal friction of an ultrafast-folding protein. *Proc Natl Acad Sci USA* 47:18320–18325.
64. Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. *Chem Rev* 106:1917–1930.
65. Zhou HX (2004) Polymer models of protein stability, folding, and interactions. *Biochemistry* 43:2141–2154.
66. Cheung MS, Thirumalai D (2006) Nanopore-protein interactions dramatically alter stability and yield of the native state in restricted spaces. *J Mol Biol* 357:632–643.
67. Elad N, et al. (2007) Topologies of a substrate protein bound to the chaperonin GroEL. *Mol Cell* 26:415–426.
68. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR (2009) Chaperonin complex with a newly folded protein encapsulated in the folding chamber. *Nature* 457:107–110.
69. Mayhew M, et al. (1996) Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* 379:420–426.
70. Martin J, Hartl FU (1997) The effect of macromolecular crowding on chaperonin-mediated protein folding. *Proc Natl Acad Sci USA* 94:1107–1112.
71. Ellis RJ (2001) Molecular chaperones: inside and outside the Anfinsen cage. *Curr Biol* 11:R1038–1040.
72. Horwich AL, Burston SG, Rye HS, Weissman JS, Fenton WA (1998) Construction of single-ring and two-ring hybrid versions of bacterial chaperonin GroEL. *Methods Enzymol* 290:114–116.
73. Müller BK, Zaychikov E, Bräuchle C, Lamb DC (2005) Pulsed interleaved excitation. *Biophys J* 89:3508–3522.

Supporting information Hofmann *et al.*

Materials and Methods

Protein preparation. GroES was expressed in BL21DE3 cells. After cell disruption in 50 mM TrisHCl, 1 mM EDTA, pH 7.5 (buffer A) and centrifugation for 15 min at 4 °C (21000 rpm, F21S50 rotor, Herolab), the supernatant was filtered and loaded on a Q-sepharose column (GE Healthcare Biosciences, Piscataway, NJ) equilibrated with buffer A (1). The column was washed with 500 ml of buffer A before starting a gradient from 0 to 1 M NaCl at a flow rate of 4 ml/min. Fractions containing GroES were combined. After heat precipitation for 20 min at 80°C in the presence of 10 mM EDTA (2), GroES was 80 % pure (according to SDS-PAGE). The remaining impurities were removed by size exclusion chromatography (Sephadex S75 26/60, GE Healthcare Biosciences, Piscataway, NJ) under denaturing conditions (50 mM TrisHCl, 1.5 M GdmCl, 1 mM EDTA, pH 7.5). Finally, the protein was refolded by dialyzing twice against 1 liter of buffer A, frozen in liquid nitrogen, and stored at -80°C. GroEL-SR1 was expressed and purified essentially as described by Horwich *et al.* (3) and stored at -80 °C as a precipitate in 2.7 M ammonium sulfate. Cysteine variants of rhodanese (E77C, D102C, K135C, K174C, D219C, K236C, E285C, E77C/K135C, K135C/K174C, K236C/E285C) were prepared and labeled with Alexa Fluor 488 and Alexa Fluor 594 (Förster radius of 5.4 nm (4-6)) as described previously (6).

We have taken the following measures to minimize the risk of perturbing the refolding kinetics by fluorophore labeling. [1] We selected chromophores with good solubility in water, which showed little or no influence on protein stability and dynamics in previous experiments (4, 7-12). [2] Previous size exclusion chromatography experiments showed no binding of the free dyes, labeled peptides, and small proteins to GroEL (6). Considering the very tight binding of rhodanese to GroEL, it seems improbable that the fluorophores have a strong effect on the interactions with the chaperone. [3] The remarkable promiscuity of GroEL/ES for different substrate proteins suggests that the chaperone will not be particularly sensitive to protein modifications. [4] We are directly comparing the folding kinetics of identically labeled proteins in free solution and inside the chaperonin cage.

Preparation of rhodanese-chaperone complexes. The ammonium sulfate precipitate of SR1 was resolubilized at 1 µM SR1 (heptamer) in folding buffer (50 mM TrisHCl (Roth), 10 mM $MgCl_2$ (Roth), 5 mM KCl (Roth), 100 mM 2-mercaptoethanol (Fluka Ultra), 0.001 % Tween 20 (Pierce), pH 7.5). Binding of rhodanese to SR1 was achieved by manually mixing 10 µl of 0.6 to 2 µM Rhodanese unfolded in 5 M GdmCl, 50 mM TrisHCl, 10 mM $MgCl_2$, 5 mM KCl, 100 mM 2-mercaptoethanol, 0.001 % Tween 20, pH 7.5 with 190 µl of a 1 µM solution of SR1 (heptamer). The complex was purified by size exclusion chromatography on a TSK 5000 PWXL column (TOSOH Bioscience) using fluorescence detection. In the presence of 1 µM GroES and 2 mM ATP, the complexes of SR1-encapsulated rhodanese were stable for more than 100 min (Fig. S1).

Refolding experiments by manual mixing. Binary complexes of SR1 and rhodanese were prepared as described above. Chaperone-mediated refolding experiments were performed in folding buffer; they were initiated by the addition of GroES (final concentration 1 µM) and ATP (final concentration of 2 mM) to the SR1-rhodanese complexes. The spontaneous refolding reaction was observed by diluting rhodanese unfolded in 4 M GdmCl 100-fold into folding buffer. For experiments at different volume fractions of D_2O, the SR1-rhodanese complexes and unfolded rhodanese were incubated in folding buffer at the respective D_2O/H_2O mixture for 1 h before refolding was initiated. All solvent isotope exchange experiments were performed at 27 °C.

Fluorescence anisotropy measurements. Steady-state fluorescence anisotropy measurements were performed in folding buffer in a Fluorolog 3 fluorometer (HORIBA Jobin Yvon, Germany). Native encapsulated rhodanese was prepared by incubating the purified binary SR1-rhodanese complexes (size exclusion chromatography, PWXL column, TOSOH Bioscience) for 1h with 2 mM ATP and 1 µM GroES. The excitation and emission wavelengths were 488 nm and 515 nm for donor and 590 nm and 615 nm for acceptor anisotropy measurements, respectively. The excitation and emission slits were 5 nm, and an integration time of 1 s was used. In total, 60 data points were measured, and an average anisotropy was calculated from these values (Fig. S2).

Reactivation of rhodanese using an enzymatic assay (13). Binary complexes of SR1 with the L-variant of rhodanese were prepared by incubating 2 µM SR1 with 1 µM donor- and acceptor-labeled rhodanese in 50 µl folding buffer for 2h at 37°C. The binary complex was purified by size exclusion chromatography on a TSK 5000 PWXL column (TOSOH Bioscience). The concentration of rhodanese in this binary complex was 0.1 µM (according to the fluorescence of Alexa Fluor 594) in 500 µl folding buffer. A premix containing 0.05 M KCN, 0.04 M KH_2PO_4, and 0.05 M $Na_2S_2O_3$ was prepared. To trigger refolding, 2 mM ATP and 1 µM GroES were added to 100 µl of binary complex (27°C). After different times (0, 10, 15, 20, 25, 30, 40, 60, 80 min), 10 µl of the reaction were mixed with 1 µl of 0.5 M EDTA (pH7.3) and cooled on ice. 125 µl of the premix were added to the mixture, and the enzymatic reaction was allowed to proceed for 1 minute before stopping it by the addition of 62.5 µl 15% formaldehyde. To detect the product, SCN^-, 187.5 µl of a 6 % (w/v) $Fe(NO_3)_3$ solution in 12 % HNO_3 were added. The complexes of iron(III) with SCN^- were detected by their absorbance at 460 nm. Three independent replicates were used to determine the rate constant for reactivation (Fig. S11). Because of the very low rhodanese concentrations used in this experiment, a background reaction due to spontaneous refolding of rhodanese during the dissociation of the complex on ice occurred. The background reaction was accounted for by repeating the reaction without the addition of GroES and ATP and subtracting it from the data. A single-exponential fit of the corrected data yielded a rate constant of 0.09 ± 0.04 min^{-1} (Fig. S11), within the range of rate constants (from ~0.04 min^{-1} to ~0.16 min^{-1}) reported in the literature for unlabeled rhodanese (14-17).

Limited proteolysis assay. Two singly-donor labeled variants (K135C-D and E285C-D) were used for limited proteolysis to probe the sequence of folding events in rhodanese (Fig. S10). Position K135C is located at the N-terminal end of the linker connecting the N- and C-terminal domains, whereas E285C is located in the C-terminal domain. The unfolded rhodanese variants (14 µM K135C-D and 7 µM E285C-D in 4 M GdmCl and folding buffer) were diluted (1:10) with folding buffer to a final volume of 200 µl. 100 seconds after starting the refolding reaction, 20 µl of a 20 mg/ml stock-solution of proteinase K were added. Proteolysis was stopped by mixing an aliquot of the reaction mixture (100 µl) with 10 µl 120% w/v TCA at 100 s and 300s after protease addition. The precipitate was dissolved in 50 µl SDS PAGE loading-buffer, and 10 µl were used for electrophoresis. The fluorescence of the attached Alexa Fluor 488 was measured in the gel (Fig. S10B) before staining with Coomassie brilliant blue R250 (Roth, Karlsruhe, Germany) (Fig. S10A). The fluorescence image (Fig. S10B, excitation wavelength 488 nm; Typhoon 9400, Amersham) shows a band with a mass of approximately 11 kD for variant E285C-D (fluorophore in the C-terminal domain), whereas only smaller fragments are observed for K135C-D (fluorophore in the N-terminal domain), suggesting that the C-terminal domain folds prior to the N-terminal domain. The molecular mass of the C-terminal domain (starting with Y163) is estimated to be 14 851 Da. Due to the high amount of protease used in the assay (2 mg/ml), no band of undigested rhodanese could be observed.

Single molecule measurements. Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant, Berlin, Germany) equipped with a 488 nm diode laser (Sapphire 488-100 CDRH, Coherent, Santa Clara, CA), a 80 MHz-pulsed laser (Optical Supercontinuum System SCF450-4-80MHz, Fianium, Southampton, UK), a 30 MHz-pulsed laser (Optical Supercontinuum System SCF450-4-30MHz Fianium), and an Olympus UplanApo 60x/1.20W objective. After passing through a 100 µm pinhole, sample fluorescence was separated first into parallel and perpendicular polarized light relative to the polarization of the exciting laser beam using a polarizing beam splitter cube. Afterwards, the two components were further separated into donor and acceptor components using a dichroic mirror (585DCXR, Chroma, Rockingham, VT). After passing two filters (Chroma ET525/50M, HQ650/100), each component was focused onto an avalanche photodiode (SPCM-AQR-15, PerkinElmer Optoelectronics, Vaudreuil, QC, Canada), and the arrival time of every detected photon was recorded using a

HydraHarp 400 counting card (PicoQuant, Berlin, Germany). All measurements were performed with laser powers of 70 µW to 100 µW. For dual color excitation of donor and acceptor (18, 19), the donor was excited continuously with a 488 nm diode laser (Sapphire 488-100 CDRH, Coherent, Santa Clara, CA) to maximize the excitation rate, and the acceptor was excited with picosecond pulses at a wavelength range selected by a z582/15 (Chroma) band pass filter and a pulse frequency of 30 MHz (Optical Supercontinuum Systems SCF450-4-30MHz Fianium, Southampton, UK). Successive photons detected in either channel and separated by less than 130 µs were combined into one burst. A burst was retained as a significant event if the total number of counts exceeded 20 (or 15 in case of the N-variant in the microfluidic mixer). Identified bursts were corrected for background, differences in quantum yields of donor and acceptor, the different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation as described previously (20). In addition, bursts during which acceptor photobleaching is likely to have occurred were discarded (6). A custom-built temperature-controlled sample holder employing peltier elements and a digital temperature controller (TC2812-LAB12, Cooltronic, Wil, Switzerland) with a PT100 platinum resistance temperature sensor (Minco EC AG, Wil, Switzerland) was used to adjust the temperature (21). The temperature in the confocal volume was determined via the temperature-dependent fluorescence lifetime of rhodamine B (RhB) (22) measured in a custom-built temperature-controlled ensemble time-correlated single photon counting instrument (23).

Data reduction by moving window analysis and singular value decomposition (SVD). For the moving window analysis, window sizes of Δt = 300 s (spontaneous folding of the N- and L-variant and SR1-mediated folding of N-, L-, and C-variant) or Δt = 50 s (spontaneous folding of the C-variant) were used. All bursts recorded within this window were used to obtain the following nine observables: transfer efficiency, burst duration, photon rate per burst (photon detection rate), donor fluorescence lifetime (parallel and perpendicular relative to the polarization of the excitation light), acceptor fluorescence lifetime (parallel and perpendicular to the polarization of the excitation light), donor anisotropy after donor excitation, and acceptor anisotropy after donor excitation. The observables were binned (25x25 bins) in two-dimensional histograms, with the transfer efficiency as one dimension. Bin-ranges were from E_{min} = -0.3 to E_{max} = 1.2 (transfer efficiency), t_{min} = 0 ms to t_{max} = 10 ms (burst duration), B_{min} = 0 ms^{-1} to B_{max} = 100 ms^{-1} (photon detection rate), r_{min} = -2 to r_{max} = 2 (anisotropy), and τ_{min} = 0 ns to τ_{max} = 8 ns (fluorescence lifetime). Every 2D-histogram was expressed column by column as a vector $\mathbf{a}_i(t_1)$ with 625 elements (25x25) for the first time window t_1. By arranging all sub-vectors $\mathbf{a}_i(t_1)$ in one vector $\mathbf{a}(t_1) = [\mathbf{a}_1(t_1), \mathbf{a}_2(t_1), \mathbf{a}_3(t_1),\ldots, \mathbf{a}_8(t_1)]$, the final vector (containing 5000 elements) for this time window (t_1) was obtained. The window was then shifted incrementally by $\Delta t/3$, and $\mathbf{a}(t_j)$ was calculated for the bursts resulting from every window position. For the SVD, a time $t = t_s+\Delta t/2$ was assigned to every histogram, where t_s is the start time of the corresponding window. The procedure yields an $m \times n$ data matrix, **A,** with n vectors $\mathbf{a}(t_j)$ for n time points and $m = 5000$ bins of the nine observables.

SVD decomposes **A** into three matrices (**U**, **S** and **V**) (24):

$$\mathbf{A} = \mathbf{U}\ \mathbf{S}\ \mathbf{V}^{\mathrm{T}}. \qquad [\mathrm{S1}]$$

Here, **U** is an $m \times m$ matrix of basis vectors (eigenvectors of $\mathbf{AA}^{\mathrm{T}}$), **S** is an $m \times n$ rectangular diagonal matrix whose elements give the weighting factors (or singular values) for every basis vector (the squared eigenvalues of $\mathbf{AA}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$), and $\mathbf{V}^{\mathrm{T}}$ is a transposed $n \times n$ matrix of amplitude vectors (eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$), describing the time course of the corresponding basis vectors. The number of nonzero diagonal elements of **S** that are necessary to reconstruct the data matrix from **U**, **S,** and **V** is an estimate of the number of distinguishable molecular species involved in the reaction mechanism. The two-dimensional representations of the first three basis vectors and the first three amplitude vectors for the spontaneous and SR1-mediated folding reaction of the N-variant, L-variant and C-variant are shown in Fig. S6, S7, S8 and S9, respectively. Raw data showing the first histograms immediately after starting the refolding reaction and the last histogram of the refolding reaction are shown for all three variants in Figs. S3, S4 and S5.

Microfluidic mixing experiments. For rapid mixing experiments, microfluidic mixers fabricated by replica molding in polydimethylsiloxane (PDMS) were used. For detecting the GroES-ATP-mediated encapsulation reaction of the SR1-bound rhodanese variants, the binary rhodanese-SR1 complex was mixed at a ratio of 1:5.7 with 2.4 µM GroES and 2.4 mM ATP, resulting in final concentrations of 2 µM GroES and 2 mM ATP. 0.01 % Tween 20 were included to prevent non-specific interactions of the chaperone-substrate complexes with the PDMS surfaces. Measurements were taken by placing the confocal volume at positions 75 µm (63 ms), 100 µm (94 ms), 200 µm (168 ms), 300 µm (252 ms), 600 µm (504 ms and 1.01 s), 900 µm (1.51 s), 1200 µm (2.02 s), and 1500 µm (2.52 s) downstream of the mixing region. To determine the transfer efficiency histogram at $t = 0$, the binary rhodanese-chaperone complex was measured 50 µm after the mixing region, without ATP and GroES in the side channels (Ch1 and Ch3, see Fig. 5A). The experiments were performed with pressures of 13.8 kPa (2.0 psi) applied to all channels for measurements from 50 µm (without GroES and ATP, 0 ms) to 600 µm (504 ms), and with 6.9 kPa (1.0 psi) for all measurements from 600 µm (1.01 s) to 1500 µm (2.52 s). The calculated flow velocities of 1.2 mm/s (13.8 kPa) and 0.6 mm/s (6.9 kPa) in the observation channel (Ch4) were used to convert distances to times as described by Pfeil *et al.* (25). The calculated velocities were confirmed by analyzing the donor-acceptor fluorescence intensity cross-correlation functions (26).

References

1. Quaite-Randall E & Joachimiak A (2000) Purification of GroES from an Overproducing E.coli Strain. *Methods Mol Biol* 140:41-49.
2. Kamireddi M, Eisenstein E, & Reddy P (1997) Stable expression and rapid purification of Escherichia coli GroEL and GroES chaperonins. *Protein Expr Purif* 11:47-52.
3. Horwich AL, Burston SG, Rye HS, Weissman JS, & Fenton WA (1998) Construction of single-ring and two-ring hybrid versions of bacterial chaperonin GroEL. *Methods Enzymol* 290:114-116.
4. Hoffmann A, Kane A, Nettels D, Hertzog DE, Baumgärtel P, *et al.* (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105-110.
5. Nettels D, Gopich IV, Hoffmann A, & Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655-2660.
6. Hillger F, Hänni D, Nettels D, Geister S, Grandin M, *et al.* (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew Chem Int Ed* 47:6184-6188.
7. Schuler B, Lipman EA, & Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743-747.
8. Sherman E & Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539-11543.
9. Hofmann H, Golbik RP, Ott M, Hübner CG, & Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* 376:597-605.
10. Tezuka-Kawakami T, Gell C, Brockwell DJ, Radford SE, & Smith DA (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys. J.* 91:L42-L44.
11. Merchant KA, Best RB, Louis JM, Gopich IV, & Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci USA* 104:1528-1533.
12. Margittai M, Widengren J, Schweinberger E, Schroder GF, Felekyan S, *et al.* (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Natl Acad Sci USA* 100:15516-15521.
13. Tandon S & Horowitz PM (1989) Reversible folding of rhodanese. Presence of intermediate(s) at equilibrium. *J Biol Chem* 264:9859-9866.
14. Chaudhry C, Farr GW, Todd MJ, Rye HS, Brunger AT, *et al.* (2003) Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *EMBO J* 22:4877-4887.
15. Weissman JS, Rye HS, Fenton WA, Beechem JM, & Horwich AL (1996) Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* 84:481-490.
16. Tang Y, Chang H, Chakraborty K, Hartl F, & Hayer-Hartl M (2008) Essential role of the chaperonin folding compartment in vivo. *EMBO J.*

17. Farr GW, Fenton WA, & Horwich AL (2007) Perturbed ATPase activity and not "close confinement" of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. *Proc Natl Acad Sci USA* 104:5342-5347.
18. Müller BK, Zaychikov E, Bräuchle C, & Lamb DC (2005) Pulsed interleaved excitation. *Biophys. J.* 89:3508-3522.
19. Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, *et al.* (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci USA* 101:8936-8941.
20. Schuler B (2007) Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol Biol* 350:115-138.
21. Nettels D, Müller-Späth S, Küster F, Hofmann H, Haenni D, *et al.* (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci USA*.
22. Benninger RK, Koç Y, Hofmann O, Requejo-Isidro J, Neil MA, *et al.* (2006) Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal Chem* 78:2272-2278.
23. Nettels D, Hoffmann A, & Schuler B (2008) Unfolded Protein and Peptide Dynamics Investigated with Single-Molecule FRET and Correlation Spectroscopy from Picoseconds to Seconds. *J Phys Chem B*.
24. Henry E & Hofrichter J (1992) Singular value decomposition - Application to analysis of experimental data. *Methods Enzymol* 210:129-192.
25. Pfeil SH, Wickersham CE, Hoffmann A, & Lipman EA (2009) A microfluidic mixing system for single-molecule measurements. *Rev Sci Instrum* 80:055105.
26. Gösch M, Blom H, Holm J, & Rigler R (2000) Hydrodynamic flow profiling in microchannel structures by single molecule fluorescence spectroscopy. *Anal Chem* 72:3260-3265.
27. Hillger F, Nettels D, Dorsch S, & Schuler B (2007) Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc* 17:759-765.

Supporting Figures

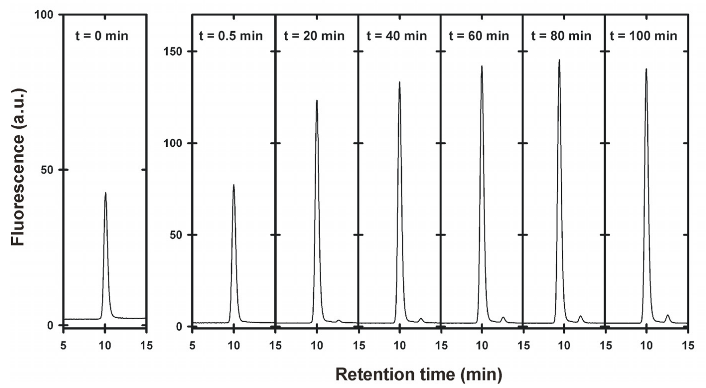


Fig. S1. Stability of the SR1-rhodanese-GroES complex. Elution profiles of analytical size exclusion chromatography of the rhodanese-SR1 complex (using singly acceptor-labeled rhodanese K135C) are shown at different times *t* after starting the refolding reaction by adding 1 μM GroES and 2 mM ATP. A TSK 5000 PWXL column (TOSOH Bioscience) with a flow rate of 0.8 ml/min and fluorescence detection was used (excitation at 585 nm, emission at 610 nm). The profile at *t* = 0 min is the size exclusion chromatography run before adding GroES and ATP. The peak at a retention time of 10 min corresponds to the chaperonin-rhodanese complex. Only very small amounts of free rhodanese (peak at 12.5 min) form over the course of the experiment. The increase in fluorescence of the peak corresponding to the chaperonin-rhodanese complex reflects the folding reaction inside the chaperonin cavity. The experiment was performed in folding buffer at room temperature.

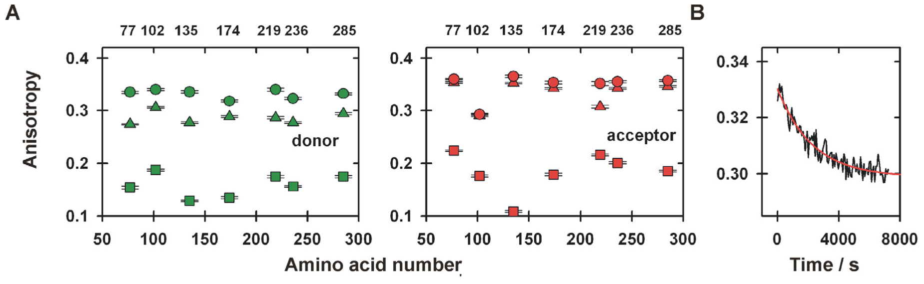


Fig. S2. Steady-state anisotropy measurements. (**A**) Single-cysteine variants of rhodanese were labeled either with the donor (green) or with the acceptor dye (red). Anisotropy values were recorded for free rhodanese (squares), binary rhodanese-SR1 complex (circles), and the encapsulated native rhodanese (triangles). (**B**) Kinetics of the change in donor anisotropy of the singly donor-labeled rhodanese variant K174C after addition of 1 µM GroES and 2 mM ATP at 22°C. The single exponential fit (solid line) yields a rate constant of $(4 \pm 1) \cdot 10^{-4}$ s^{-1}. The good agreement with the rate constants obtained from the single-molecule FRET experiments on doubly labeled protein [$(3.2 \pm 1.3) \cdot 10^{-4}$ s^{-1} for the N-variant and $(3.0 \pm 0.5) \cdot 10^{-4}$ s^{-1} for the L-variant at 22°C] supports the accuracy of the single molecule results and provides some indication that dye labeling does not affect the kinetics significantly (at least for AlexaFluor594, which is the larger chromophore). All measurements were performed in folding buffer.

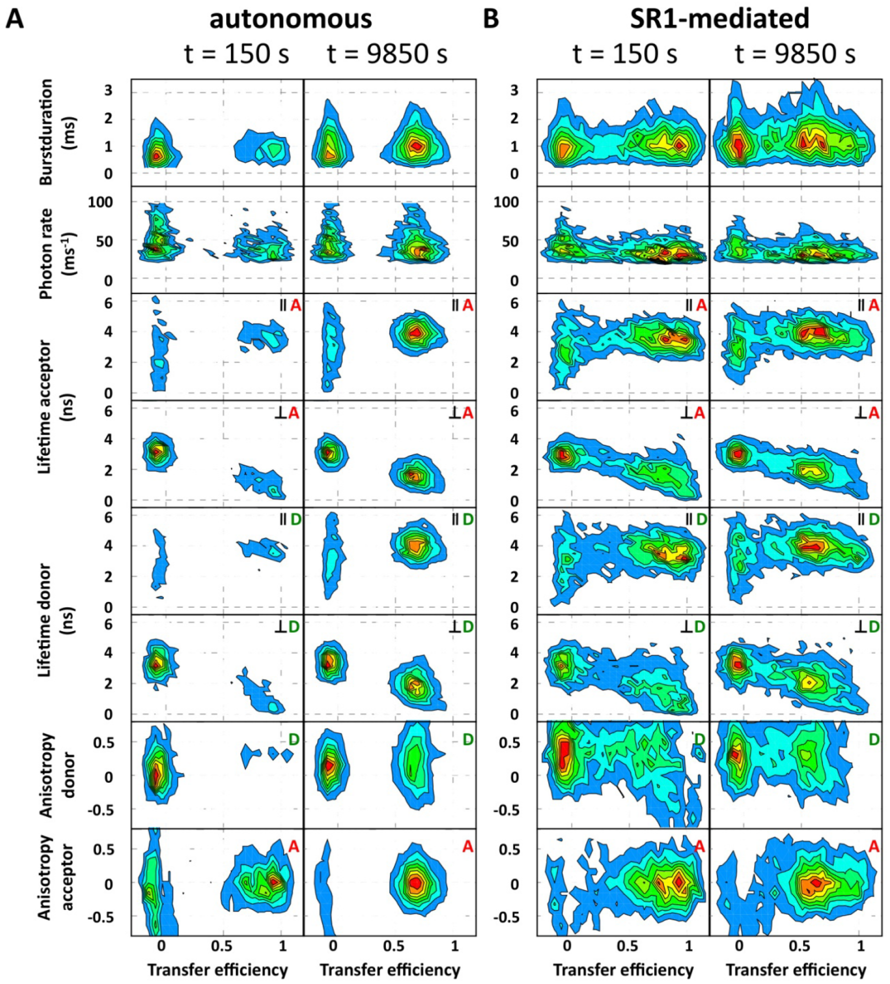


Fig. S3. 2D-Histograms for the autonomous **(A)** and SR1-mediated **(B)** folding reaction of the N-variant at the earliest (left) and latest (right) time point after starting the refolding reaction at 24°C.

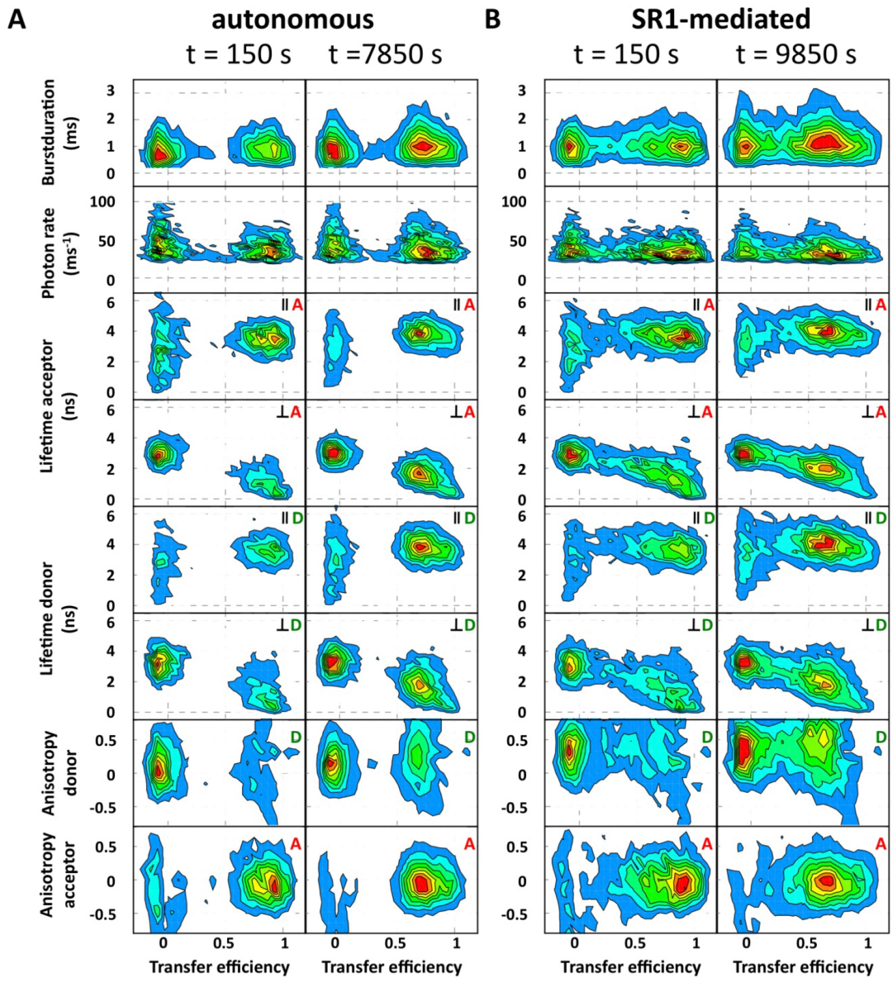


Fig. S4. 2D-Histograms for the autonomous **(A)** and SR1-mediated **(B)** folding reaction of the L-variant at the earliest (left) and latest (right) time point after starting the refolding reaction at 24°C.

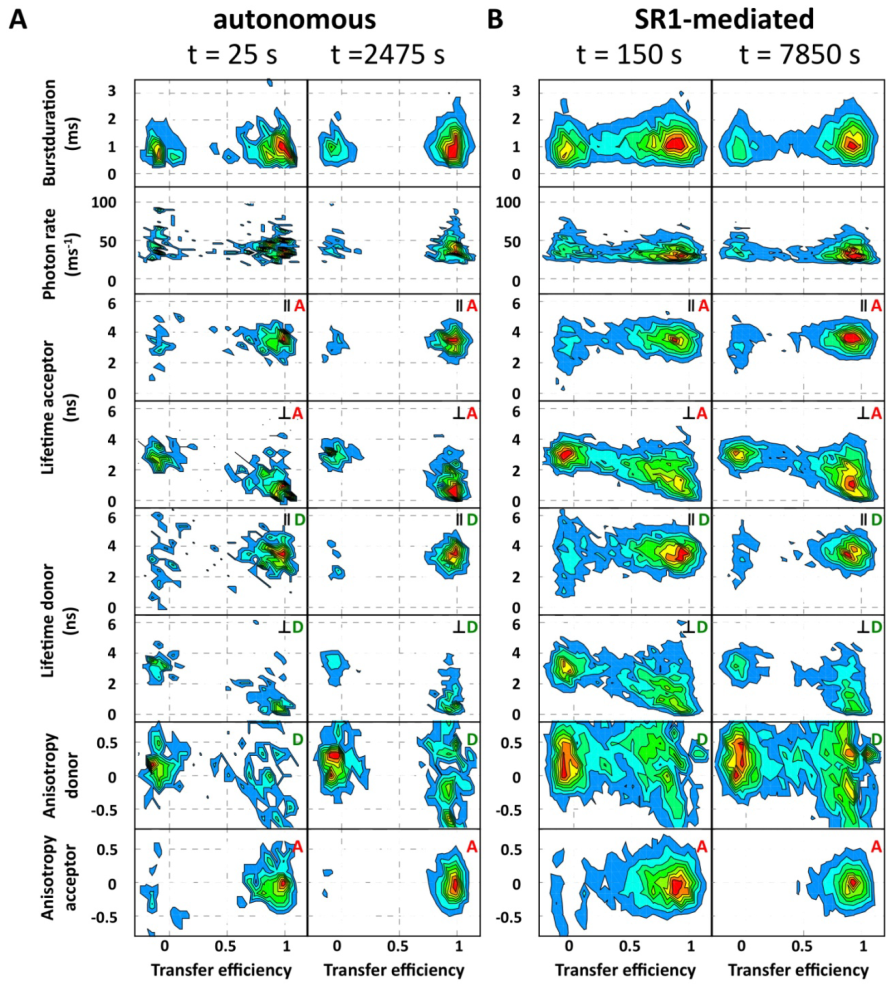


Fig. S5. 2D-Histograms for the autonomous **(A)** and SR1-mediated **(B)** folding reaction of the C-variant at the earliest (left) and latest (right) time point after starting the refolding reaction at 24°C.

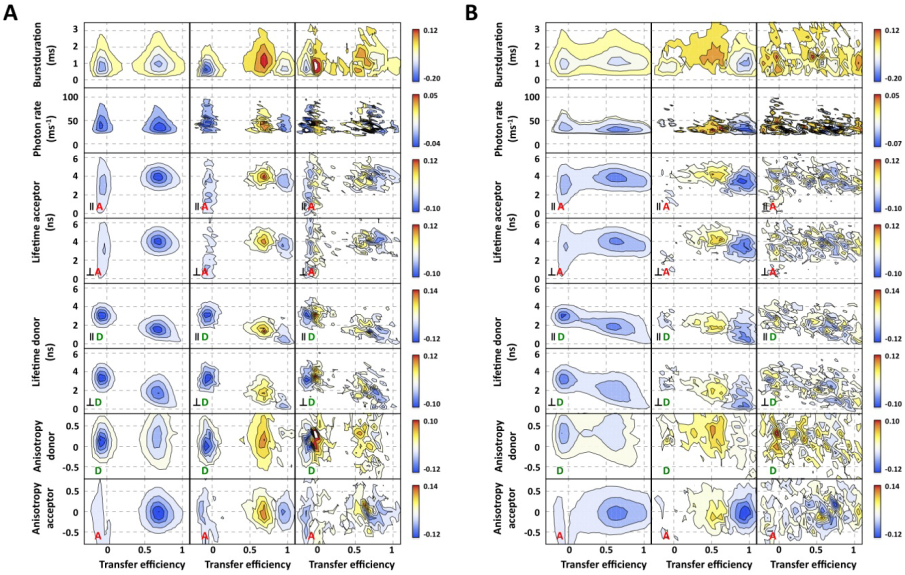


Fig. S6. Two-dimensional basis vectors of the multi-dimensional SVD for the spontaneous **(A)** and SR1-mediated **(B)** folding reaction of the N-variant at 24°C. The basis vectors indicate the positions of changes in the histograms of the corresponding observables (from top to bottom) and are ordered according to their singular values (from left to right). Basis vector 1 (left panel) describes an increase in the number of molecules during refolding. Basis vector 2 (middle panel) describes the conversion of the signature of non-native molecules (blue) to that of the native state (yellow and red). The larger noise level of basis vector 3 and the small corresponding singular values (see Fig. S9) for both the spontaneous (A) and the SR1-mediated folding reaction (B) indicate that the folding reaction is dominated by two distinguishable molecular species. The color code reflects the absolute SVD amplitude (see color scale).

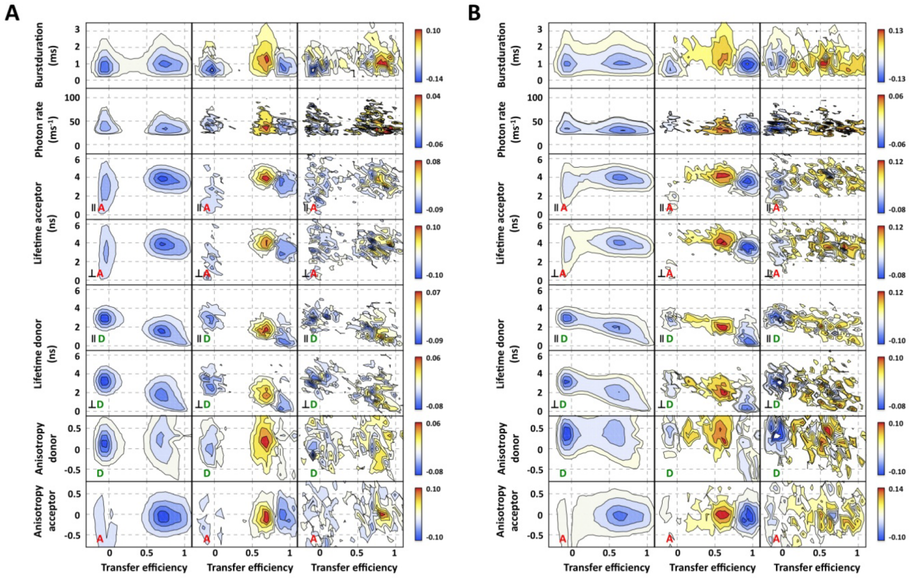


Fig. S7. Two-dimensional basis vectors of the multi-dimensional SVD for the spontaneous **(A)** and SR1-mediated **(B)** folding reaction of the L-variant at 24°C. The basis vectors indicate the positions of changes in the histograms of the corresponding observables (from top to bottom) and are ordered according to their singular values (from left to right). Basis vector 1 (left panel) describes an increase in the number of molecules during refolding. Basis vector 2 (middle panel) describes the conversion of the signature of non-native molecules (blue) to that of the native state (yellow and red). The larger noise level of basis vector 3 and the small corresponding singular values (see Fig. S9) for both the spontaneous (A) and the SR1-mediated folding reaction (B) indicate that the folding reaction is dominated by two distinguishable molecular species. The color code reflects the absolute SVD amplitude (see color scale).

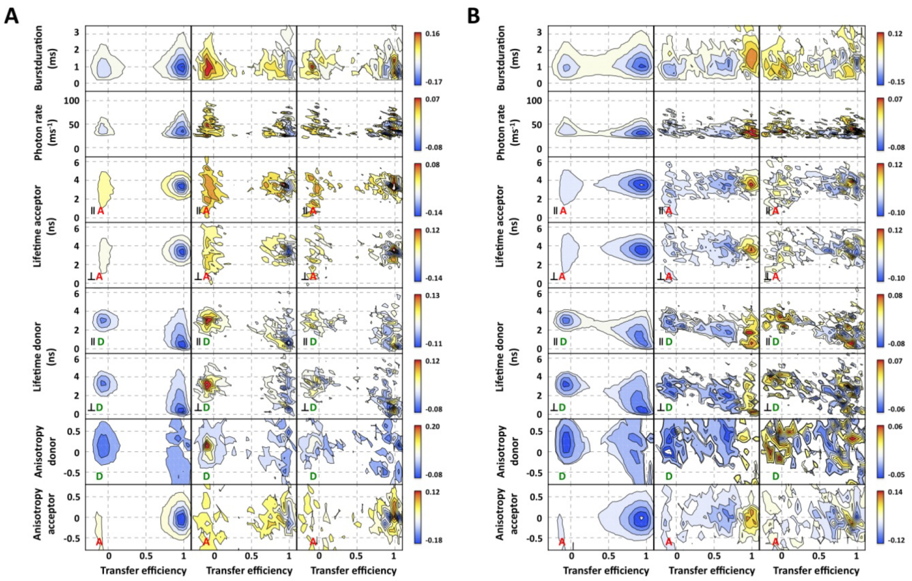


Fig. S8. Two-dimensional basis vectors of the multi-dimensional SVD for the spontaneous **(A)** and SR1-mediated **(B)** folding reaction of the C-variant at 24°C. The basis vectors indicate the position of a change in the histogram of the corresponding observable (from top to bottom) and are ordered according to their singular values (from left to right). Basis vector 1 (left panel) describes an increase in the number of molecules during refolding. Basis vector 2 (middle panel) describes the conversion of the signature of non-native molecules (blue) to that of the native state (yellow and red). The larger noise level of basis vector 3 and the small corresponding singular values (see Fig. S9) for both the spontaneous (A) and the SR1-mediated folding reaction (B) indicate that the folding reaction is dominated by two distinguishable molecular species. The color code reflects the absolute SVD amplitude (see color scale).

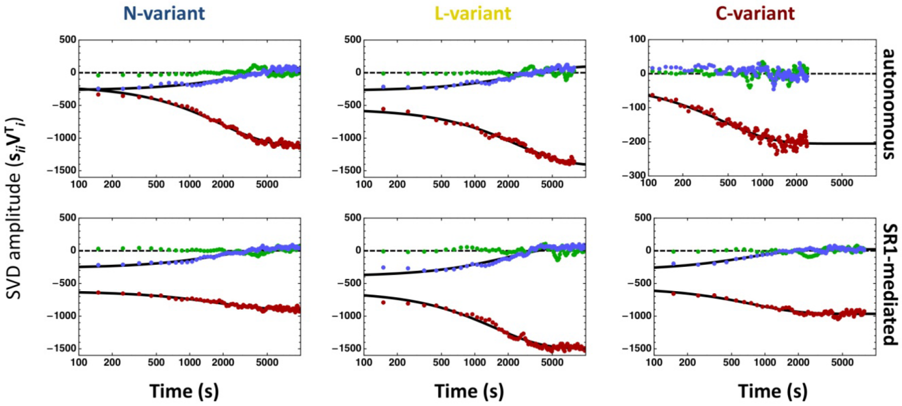


Fig. S9. First (red), second (blue), and third (green) amplitude vector weighted by their corresponding singular values (s_{ii}) of the multi-dimensional SVD for the spontaneous (upper panels) and SR1-mediated (lower panels) folding reaction of the N-, L-, and C-variants at 24°C. The first two amplitude vectors dominate the observed signal change.

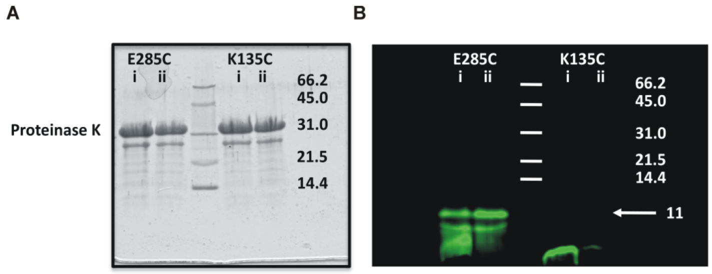


Fig. S10. Limited Proteolysis of singly-labeled rhodanese (K135C-D and E285C-D) indicates that the C-terminal domain folds prior to the N-terminal domain. **(A)** Coomassie stained SDS-Polyacrylamide gel (17%) of the proteolysis reaction mixture stopped at 100 s (i) and 300 s (ii) after addition of 2 mg/ml proteinase K. **(B)** Same gel, but scanned with an excitation wavelength of 488 nm.

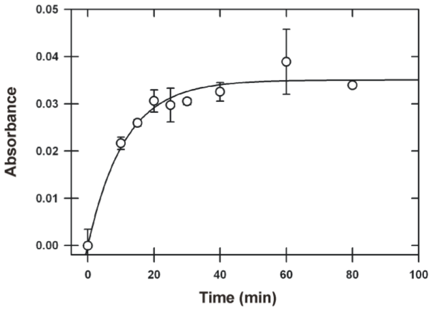


Fig. S11. Reactivation of doubly labeled rhodanese K135C/K174C by SR1. Enzymatic test as described in the supporting text. Error bars show standard errors of the means resulting from 3 independent measurements. A single-exponential fit yields a rate constant of 0.09 ± 0.04 min^{-1}, within the range of rate constants (from ~0.04 min^{-1} to ~0.16 min^{-1}) reported in the literature (14-17). The reactivation kinetics agree well with our single molecule results, especially considering the spread of previously reported values, the lack of temperature control for some of them, and the slow GroEL/ES dissociation step on ice required prior to the enzymatic test (13).

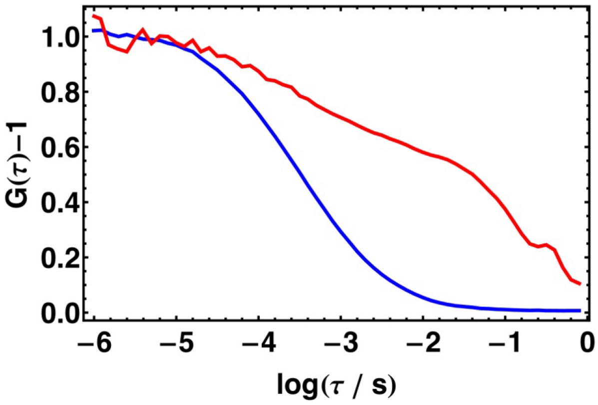


Fig. S12. Monitoring aggregation in single molecule experiments with fluorescence correlation spectroscopy. Under the experimental conditions and picomolar protein concentrations used in this work, no aggregation of rhodanese occurred, as illustrated by the normalized donor-acceptor fluorescence intensity cross-correlation function of the L-variant after initiation of refolding by manual mixing at 24°C (blue). The signatures of aggregated rhodanese previously observed in fluorescence lifetimes and burst size distributions (27) were also absent under these conditions. For comparison, a normalized cross-correlation function of rhodanese L-variant in the presence of an excess of unlabeled wildtype rhodanese is shown in red. The sample containing 4 nM labeled L-variant and 3 μM rhodanese wildtype was incubated for 2 minutes in 3 M urea. After dilution 1:50 into native conditions, the single molecule fluorescence time trace was recorded and the cross-correlation function calculated. The presence of aggregates leads to a heterogeneous distribution of translational diffusion times through the confocal volume and a correspondingly slower and in some cases even non-monotonic decay of the correlation function.

5 Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach

M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, and H. Grubmüller.

Contributions of Dominik Hänni:

Martin Hoefling performed MD simulations on fluorescently labeled polyprolines where he found that sticking of the donor dye to the polyproline leads to a change in the FRET behavior. Having the experimental tools at hand, I decided to search for such an effect. For this, suitably labeled polyproline samples were highly purified to avoid contributions by donor-only molecules. These purified peptides were measured in various solvent conditions to screen possible interactions between the fluorescent dye and the peptide. An extensive analysis of the experimental data revealed no experimental indication for such a sticking behavior. Additionally, the experimentally determined transfer efficiency histograms were used for a direct comparison to the simulated transfer efficiency histograms, showing a good agreement between simulated and experimental single-molecule data.

OPEN ACCESS Freely available online PLoS one

Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach

Martin Hoefling[1], **Nicola Lima**[1], **Dominik Haenni**[2], **Claus A. M. Seidel**[3], **Benjamin Schuler**[2], **Helmut Grubmüller**[1]*

1 Theoretical and Computational Biophysics Department, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, **2** Department of Biochemistry, University of Zurich, Zurich, Switzerland, **3** Institute of Molecular Physical Chemistry (MPC), Heinrich Heine University, Düsseldorf, Germany

Abstract

Förster Resonance Energy Transfer (FRET) experiments probe molecular distances via distance dependent energy transfer from an excited donor dye to an acceptor dye. Single molecule experiments not only probe average distances, but also distance distributions or even fluctuations, and thus provide a powerful tool to study biomolecular structure and dynamics. However, the measured energy transfer efficiency depends not only on the distance between the dyes, but also on their mutual orientation, which is typically inaccessible to experiments. Thus, assumptions on the orientation distributions and averages are usually made, limiting the accuracy of the distance distributions extracted from FRET experiments. Here, we demonstrate that by combining single molecule FRET experiments with the mutual dye orientation statistics obtained from Molecular Dynamics (MD) simulations, improved estimates of distances and distributions are obtained. From the simulated time-dependent mutual orientations, FRET efficiencies are calculated and the full statistics of individual photon absorption, energy transfer, and photon emission events is obtained from subsequent Monte Carlo (MC) simulations of the FRET kinetics. All recorded emission events are collected to bursts from which efficiency distributions are calculated in close resemblance to the actual FRET experiment, taking shot noise fully into account. Using polyproline chains with attached Alexa 488 and Alexa 594 dyes as a test system, we demonstrate the feasibility of this approach by direct comparison to experimental data. We identified *cis*-isomers and different static local environments as sources of the experimentally observed heterogeneity. Reconstructions of distance distributions from experimental data at different levels of theory demonstrate how the respective underlying assumptions and approximations affect the obtained accuracy. Our results show that dye fluctuations obtained from MD simulations, combined with MC single photon kinetics, provide a versatile tool to improve the accuracy of distance distributions that can be extracted from measured single molecule FRET efficiencies.

Citation: Hoefling M, Lima N, Haenni D, Seidel CAM, Schuler B, et al. (2011) Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. PLoS ONE 6(5): e19791. doi:10.1371/journal.pone.0019791

Editor: Jerome Mathe, Université d'Evry val d'Essonne, France

Received January 11, 2011; **Accepted** April 5, 2011; **Published** May 24, 2011

Funding: This work has been supported by the German Israeli Foundation grant (Agreement No. 1000.89.9/2008) (to MH), Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) (to MH), the Swiss National Science Foundation (to BS) and the Volkswagen Stiftung (to HG, CS and BS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hgrubmu@gwdg.de

Introduction

Since the development of the Resonance Energy Transfer theory by Förster (FRET) in the late forties [1], and the definition of this technique as a "spectroscopic ruler" in biological systems by Stryer and Haugland [2], single molecule detection [3–5] and time-resolved experiments [6] have opened up a new window to probe inter- and intramolecular distances and motions. In a typical experiment, donor molecules are excited by a laser pulse, and part of the excitation energy is transferred to nearby acceptor molecules. The transfer efficiency

$$E=\frac{I_{\mathrm{A}}}{I_{\mathrm{D}}+I_{\mathrm{A}}} \tag{1}$$

is measured via the donor fluorescence intensity I_{D} and the acceptor fluorescence intensity I_{A}. Among other factors, E depends on the distance R between the donor and the acceptor fluorophores, as well as on the mutual orientation of their respective transition dipole moments. After orientational averaging, the distance dependency is described by Förster's approximation,

$$E=\frac{1}{1+\left(\frac{R}{R_0}\right)^6}\quad, \tag{2}$$

where R_0 is the so-called Förster radius which denotes the distance at which 50% of the donor excitation is transferred to the acceptor molecule.

This relation is widely used to monitor structural changes in biomolecules via FRET efficiency measurements [2,7]. To that aim, donor and acceptor fluorophores are covalently attached to specific sites of the macromolecule of interest. Taking into account the flexibility of the fluorophores and their linkers, the measured intensities provide information on the mutual distance of these specific sites [8–11]. The use of multiple dye pairs allows for triangulation of biomolecules, which provides three-dimensional structural information [10,12–16].

In single molecule setups, distributions and distance fluctuations of individual molecules are accessible [4,17–19]. If the scatter of the observed efficiency distributions in these experiments is broader than the expected shot noise, distance distributions can be estimated [20]. For distance changes in the biomolecule, which are slow compared to the burst duration, time resolved information is then accessible [21,22]. By recording millisecond fluorescence bursts while the molecules diffuses through a confocal laser volume, conformational motions in the same time scale have been resolved [21,23,24].

FRET spectroscopy has proven particularly successful in situations where the mutual orientation distribution of the transition dipole moments can be considered isotropic and uncorrelated. Examples are freely diffusing dyes, or dyes attached to flexible and solvent-exposed parts of a protein [18] or nucleic acids [10,11]. In this case, orientational averaging gives rise to the well-known orientation factor $\kappa^2 = 2/3$, which is by convention included within the Förster radius R_0 [7]. In contrast to this average κ^2, the instantaneous orientation factor $\kappa^2(t)$ can assume values in the range of 0 to 4.

Particularly when triangulating biomolecules, however, the dye motion is often far from isotropic due to steric restrictions set by the biomolecule, as well as due to electrostatic or hydrophobic interactions between the dye and the protein surface [25–30]. Since the mutual dye orientation is typically inaccessible to experiments, the $\kappa^2 = 2/3$ approximation provides only qualitative insights, unless the free and rapid reorientation of the dyes is commonly verified by fluorescence anisotropy measurements [31]. For this reason, efficiency distributions rather than distances are often reported.

The orientational dynamics uncertainty of fluorophores has been addressed via several routes. Empirical, semi-empirical, and theoretical models [32–36] for the orientational factor have been developed, assuming that the dynamics of the dyes can indeed be described by a time average. Recent computer simulations [37,38] have suggested that the mutual dye orientation can be highly anisotropic, with κ^2-values deviating markedly from 2/3 (0.24–1.02 [38]; 0.71–2.81 [37]). R_0 has been refined through fluorescence quenching measurements of multiple fluorophores [39].

Despite these efforts, three main problems remain. First, the assumption of an isotropic dye orientation distribution is invalid or difficult to establish in most cases [40,41]. Second, possible correlations between the distance and dye orientation distribution are neglected in the above treatments [38]. Third, the orientational sampling during individual bursts may be incomplete, in which case the dye distribution relevant for the observed efficiency depends on the duration of the bursts. In all three cases, applying an average κ^2 – as opposed to the κ^2 of instantaneous and time-dependent Förster transfer rate coefficients – leads to an additional broadening of the efficiency distribution [25], and biased distance distributions are obtained.

To overcome these limitations, we have developed an approach that combines molecular dynamics (MD) simulations of a dye-labeled biomolecule in solution with Monte Carlo (MC) simulations of dye excitation, FRET transfer, and fluorescence decay events. This approach involves four steps.

First, extended and fully atomistic MD simulations of the solvated biomolecule, labeled with a FRET dye pair, serve to cover the biomolecular dynamics at the fluorescence decay time scales of the system. To capture structural motions that are slower than the nanoseconds time scale accessible to MD simulation, several MD trajectories are recorded starting from different isomers and combined into a comprehensive ensemble using appropriate Boltzmann weights.

In the second step, time-dependent mutual dye orientations extracted from these trajectories are recorded. These orientations are then used to derive time-dependent instantaneous resonance energy transfer rate coefficients $k_\mathrm{T}(t)$. Within a short time interval Δt, these rate coefficients specify the probability $p_\mathrm{T}(t) = \Delta t \cdot k_\mathrm{T}(t)$ that a FRET transfer event takes place, for each instant of time.

In the third step, using $p_\mathrm{T}(t)$, a large number of MC runs is carried out to simulate and collect many individual photon absorption and excitation, FRET transfer, and emission events. For each photon absorption event, an instant of the trajectories is chosen randomly, and the probabilities are propagated appropriately until a photon emission or radiationless decay event occurs. After averaging over sufficiently many events, fluorescence intensities I_D and I_A are calculated. The numbers of photos recorded from the donor and the acceptor dyes, respectively, finally determine an average FRET efficiency value E. Similar approaches using dye conformations from simulations have been proposed recently [42–45].

To mimic single molecule FRET (smFRET) experiments, in a fourth step the emitted photons are collected into bursts according to the experimental photon burst size distribution (BSD). The efficiency in each burst is then calculated, and efficiency histograms are obtained, similar to single molecule experiments. By construction, this procedure takes shot noise accurately into account.

This hybrid simulation approach will enable one to calculate efficiency distributions that can be directly compared to measured efficiency distributions. Vice versa, we will develop a systematic approach to reconstruct distance distributions by combining the dye orientation and photon statistics at hand with measured efficiency distributions.

Here we apply this approach to a polyproline 15, 20, and 30-mer [46] with two FRET dyes (Alexa 488 and 594, Fig. 1) attached to both termini [2,31,45](Fig. 2A). As dye-labeled polyproline chains have been widely used as "rigid rods" to test the validity of the approximations underlying Förster's theory, and to gauge the Förster radius of several of FRET pairs in different environments [2,31,45], much of the current understanding relies on the particular properties of these systems. Initially assumed to be quite rigid, all-trans polyproline helices were used in the definition of FRET as a "spectroscopic ruler" [2]. This assumption was challenged quite early [47,48], suggesting that polyproline chains exhibit a substantial degree of flexibility [49]. The issue is still not fully resolved.

For these reasons, polyproline flexibility has been revisited recently by performing single molecule FRET recordings [31,50] and simulations [45] on these molecules. Indeed, unexpectedly broad efficiency distributions were seen, suggesting substantial structural heterogeneity. A detailed analysis of single molecule data showed the heterogeneity that persists on time scales greater than 10 ns [50]. Recent NMR experiments [45] pointed to a considerable population of *cis*-isomers within all-*trans* polyproline helices, which might contribute further to the structural flexibility and heterogeneity of polyprolines. These findings put the suitability of these molecules as "rigid rods" in question, and the unexpected complexity of their dynamics requires a detailed study of the structural ensemble in solution at room temperature.

Here we attempt a comprehensive characterization of the polyproline structural heterogeneity by combining atomistic

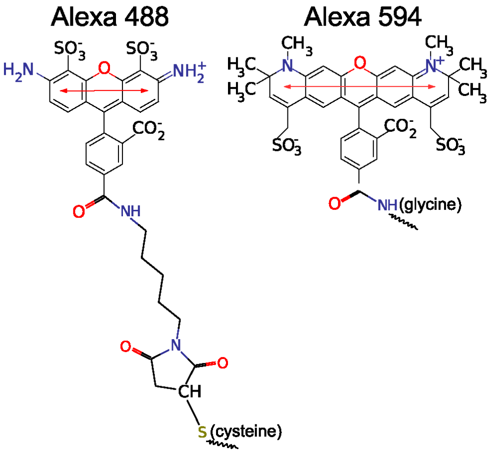


Figure 1. Dye and Linker Structures. Structure and transition dipole moments of Alexa 488 and Alexa 594. The red arrows show the orientation of the transition dipole moments. MarvinSketch was used to draw the chemical structures, Marvin 5.3.0.2 , 2010, ChemAxon (http://www.chemaxon.com).
doi:10.1371/journal.pone.0019791.g001

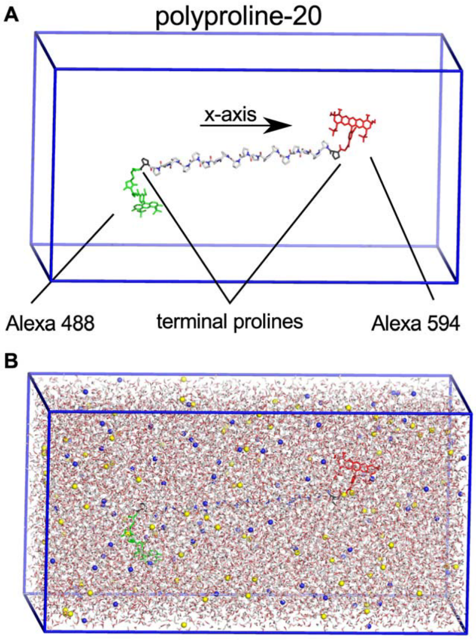


Figure 2. System Setup. (A) All-*trans* polyproline-20 molecular structure including Alexa 488 (green) and Alexa 594 (red) dyes attached by their corresponding linkers. The simulation box is shown in blue, terminal prolines used to restrain the position are depicted in black. (B) Fully solvated system is shown including Na^+ (blue) and Cl^- (yellow) ions.
doi:10.1371/journal.pone.0019791.g002

simulations with single molecule FRET data. Resting on a direct comparison of single burst efficiencies collected over many bursts, our approach is based on much fewer assumptions than the standard interpretation of FRET experiments. In particular, this approach includes κ^2 averages, on the basis of the detailed molecular dynamics of the system, and cases where the motion of the dyes is slower than the donor fluorescence decay time are readily handled. Moreover, all possible correlations between the dye movement and the distances are included, such that accurate mutual orientation distributions are obtained. Finally, the approach fully accounts for the photon count shot noise. Vice versa, comparison with experiments will enable us to test our approach. As we will demonstrate, our approach serves to combine dye orientational dynamics from MD with experimental FRET efficiency distributions at increasingly refined approximation levels.

The good agreement of distance distributions of polyproline obtained by this approach with the reference distribution suggests that this combination allows extraction of improved quantitative geometrical information from single molecule FRET experiments. By comparison with synthetic FRET data, the validity of the reconstruction will be established.

Methods

System Setup

The studied system comprises a polyproline peptide of 15, 20 or 30 proline residues [46], an amino-terminal glycine and a carboxyl-terminal cysteine residue, to which a succinimide ester and maleimide derivatives of Alexa 594 and Alexa 488 dyes [52](Fig. 1), respectively, are attached. Figure 2A shows the simulation system for the polyproline-20 [53] within a rectangular simulation box. Figure 2B depicts the box filled with explicit water molecules and 300 mM NaCl, corresponding to the ionic strength of 50 mM sodium phosphate buffer used in the experiment [31]. The number of Na^+ and Cl^- ions was chosen such as to obtain a neutral system.

In aqueous solution the most stable configuration for polyproline chains is the polyproline II (PPII) helix [53,54], characterized by dihedral angle values Φ,Ψ and Ω of -75°, 150°, and 180°, respectively [47], with the *trans*-isomer as the most favorable configuration. Nevertheless, in water a marked fraction of *cis* peptide bonds the PPII helices is observed. By NMR experiments a fraction of approximately 10% for proline at the C-terminus of the chain and 2% within the chain was measured [45], with *trans* to *cis* transition times of 10^3 to 10^4 seconds [50,55]. As this is far beyond MD time scales, separate simulations were performed for all relevant isomers, for subsequent weighted averaging. To this end, all possible isomers containing one single *cis* peptide bond were considered, i.e., 20 *cis*-trajectories for the polyproline-20 with dyes attached. Additionally, for polyproline-30, a subset of 61 isomers with *cis*-bonds at two positions was simulated.

Force Field

For water molecules, the TIP4P model was employed [56]. Force field parameters for the peptide were taken from a modified OPLS-AA force field [57] including custom parameters for the two dyes and their corresponding linkers. Alexa 488 and Alexa 594 are highly conjugated systems whose parameters are not included within the standard OPLS-AA force field. Figure 1 depicts the atomic structure of the two dyes together with the orientation of the transition dipole moments. All dye parameters (bonded and Lennard-Jones) – except for the partial charges –

were assigned via an analogy approach from similar OPLS-AA groups [58].

Because FRET occurs when the donor dye is in the excited state and the acceptor in the ground state, partial charges of these corresponding states were used in all our simulations for the dyes. The fact that the partial charges calculated for the ground and excited states differed only by a small amount suggests that the effect of this simplified treatment on the dynamics of the dyes is small. All partial charges were calculated by fitting to the electrostatic potential surfaces (EPS approach [59]) obtained from *ab-initio* B3LYP Density Functional Theory (DFT) calculations with the 6-31G* basis set. All *ab-initio* calculations were performed with the GAUSSIAN 03 program package [60]. First, for reference, the point charges for the 20 natural amino acids were calculated with B3LYP/6-31G* CHelpG population analysis to assure compatibility of the derived charges with OPLS-AA. A mean scaling factor of 0.9 was calculated by averaging the multiplicative factors of each amino-acid, which minimizes the mean square deviation between OPLS-AA and DFT charges (amino-acid scaling factors shown in Suppl. Table S1).

For the ground state of the two dyes, the same protocol was used. For the excited state, we determined charge differences with respect to the ground state for each atom in two steps. First, point charges were determined from Configuration Interaction Singlets (CIS) calculations for the first excited state using the STO-3G basis set. From these values, in a second step, point charges were subtracted, that were obtained from Hartree Fock (HF) calculations with the same STO-3G basis.

For both, ground and excited state, the charges were averaged to reflect the internal symmetry of the molecule, and scaled with the previously calculated scaling factor of 0.9. Finally, a small offset was added to all partial charges to re-establish the correct total charge of the system.

Molecular Dynamics Simulations

All MD simulations were carried out with the GROMACS 4.0.7 simulation software package [61–63]. Each proline system was energy-minimized by steepest descent to convergence. Periodic boundary conditions were applied in all three dimensions. V-Sites on hydrogens [64] were used allowing 4 fs integration time steps. After minimization, 10 ns equilibration simulations were performed. From the last 5 ns of these simulations, starting conformations for all subsequent production runs were selected at random instances (Table 1). Solvent and ions as well as the solute were separately coupled to an external temperature bath with a time constant of 0.1 ps applying the v-rescale algorithm [65,66]. The system was coupled to an isotropic pressure bath of 1 atm using the Parinello-Rahman algorithm [67] and a time constant of 1 ps. Bond lengths were constrained to their equilibrium lengths with LINCS [68]. The cut-off for Lennard-Jones interactions was set to 1 nm. Electrostatic interactions between charged groups at distances below 1 nm were calculated in direct space, while for the long-range interactions the particle-mesh-Ewald method [69] with a grid spacing of 0.12 nm and fourth order spline interpolation was used. All simulations were performed with random Maxwell-distributed starting velocities at 293 K, 303 K, and 313 K (Table 1).

Soft restraints were imposed to suppress rotation of the entire molecule in the box and thus to allow the use of a small simulation box, adapted to the shape of the molecule. To this end, the component of the difference vector perpendicular to the x-axis (Fig. 2A) between the centers of mass of the two terminal prolines was restrained to zero with a weak harmonic potential ($k = 9.744\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$, corresponding to a Boltzmann distribution of width $\sigma = 0.5\,\mathrm{nm}$). We assume that these soft restraints leave the internal dynamics of the molecule unperturbed.

Table 1. Performed molecular dynamics simulations.

Proline Length	Isomer	Temperature	Number of simulations
		K	
pro15	all-*trans*	293	10
pro15	single-*cis*	293	30
pro20	all-*trans*	293	20
pro20	all-*trans*	303	10
pro20	all-*trans*	313	10
pro20	single-*cis*	293	40
pro30	all-*trans*	293	10
pro30	single-*cis*	293	30
pro30	double-*cis*	293	61

Simulations are listed according to isomer and applied temperature. Single-*cis* simulation were carried out for all possible *cis*-isomer positions. For polyproline-30, in addition, a representative set of 61 isomers, randomly picked from the 870 possible isomers with two *cis* bonds, was simulated. All simulation lengths are 100 ns summing up to a total sampling of 22.1 μs.

doi:10.1371/journal.pone.0019791.t001

Resonance Energy Transfer Rates

All FRET efficiencies were calculated from the MD simulations using following kinetics,

$$\begin{array}{ccccc} & & \mathrm{D}+\mathrm{A} & & \mathrm{D}+\mathrm{A} \\ & & \uparrow k_{\mathrm{D_i}} & & \uparrow k_{\mathrm{A_i}} \\ \mathrm{D}+\mathrm{A}+h\nu & \rightarrow & \mathrm{D}^*+\mathrm{A} & \overset{k_{\mathrm{T}}(t)}{\rightarrow} & \mathrm{D}+\mathrm{A}^* \\ & & \downarrow k_{\mathrm{D}} & & \downarrow k_{\mathrm{A}} \\ & & \mathrm{D}+\mathrm{A}+h\nu_{\mathrm{D}} & & \mathrm{D}+\mathrm{A}+h\nu_A \end{array} \tag{3}$$

Starting after a photon adsorption event by the donor dye, this kinetics is described by

$$\dot{p}(\mathrm{D}^*+\mathrm{A}) = -\left(k_{\mathrm{D_i}}+k_{\mathrm{D}}+k_{\mathrm{T}}(t)\right)\cdot p(\mathrm{D}^*+\mathrm{A}) \text{ and} \tag{4}$$

$$\dot{p}(\mathrm{D}+\mathrm{A}^*) = k_{\mathrm{T}}(t)\cdot p(\mathrm{D}^*+\mathrm{A}) - \left(k_{\mathrm{A_i}}+k_{\mathrm{A}}\right)\cdot p(\mathrm{D}+\mathrm{A}^*). \tag{5}$$

In Eq. 3, D/D^* is the donor (Alexa 488) and A/A^* is the acceptor (Alexa 594) dye in their ground and the excited state, respectively. $h\nu, h\nu_{\mathrm{D}}$ and $h\nu_{\mathrm{A}}$ denote the exciting photon and photons emitted by the donor and the acceptor dye. The rate coefficients refer to FRET (k_T), fluorescence and internal conversion of the donor (k_{D},$k_{\mathrm{D_i}}$), and fluorescence as well as internal conversion of the acceptor dye (k_{A},$k_{\mathrm{A_i}}$).

The rate coefficients were calculated from the lifetimes τ of the dyes and their respective quantum yields Q,

$$k_{\mathrm{D}} = \frac{Q_{\mathrm{D}}}{\tau_{\mathrm{D}}},\ k_{\mathrm{A}} = \frac{Q_{\mathrm{A}}}{\tau_{\mathrm{A}}}, \tag{6}$$

$$k_{\mathrm{D_i}} = \frac{1-Q_{\mathrm{D}}}{\tau_{\mathrm{D}}},\; k_{\mathrm{A_i}} = \frac{1-Q_{\mathrm{A}}}{\tau_{\mathrm{A}}}. \tag{7}$$

For the Alexa 488 and 594 dyes attached to polyproline peptides, we used the measured lifetimes τ_{D} of 4.0 ns and $\tau_{\mathrm{A}} = 3.9$ns. To obtain photon statistics directly comparable to the experiment, the quantum yields were combined with the detector efficiencies into (relative) effective quantum yields using the correction matrix defined in Ref. [70]. In this framework, Q_{A} and Q_{D} correspond to the diagonal correction matrix elements. For the simulations, we averaged the two detector channels used in the experiment, yielding 0.77 and 1.0 for donor and acceptor effective quantum yields, respectively. Crosstalk, direct acceptor excitation, and background were found to change the photon statistics only by a small amount and thus are neglected in our MC approach.

For the time-dependent FRET rate coefficient $k_{\mathrm{T}}(t)$, which depends on the electronic coupling between the two dyes and thus also on their mutual orientation at each instant, we used Förster's dipole approximation for the electronic coupling,

$$k_{\mathrm{T}} = \left(k_{\mathrm{D}} + k_{\mathrm{D_i}}\right)\left(\frac{R_0}{R}\right)^6. \tag{8}$$

In Eq. 8, R is the distance between the geometric center of the ring system of the acceptor and the donor dyes, and R_0 is the Förster radius (the distance of 50% excitation transfer), which is proportional to the time-dependent orientation factor κ^2,

$$R_0^6 = \frac{9(\ln 10)Q_{\mathrm{D}}J\kappa^2}{128\pi^5 n^4 N_{\mathrm{A}}} = R_{\mathrm{const}}^6 \cdot \kappa^2, \tag{9}$$

where Q_{D} is the quantum yield of the donor in the absence of the acceptor, J the spectral overlap integral (Franck Condon factor), N_{A} Avogadro's number, n the index of refraction of the solvent, and κ^2 is the time-averaged orientation factor [3,70,71]. For the pair Alexa 488 – Alexa 594, a Förster radius R_0 of 5.4 nm has been determined [7,72], based on the assumption of isotropic dye orientations i.e., $\kappa^2 = 2/3$. To describe time-dependent Förster transfer, R_0^6 in Eq. 8 is therefore replaced by $R_{\mathrm{const}}^6 \kappa^2(t)$, with $R_{\mathrm{const}} = 5.4\,\mathrm{nm}/\sqrt[6]{\frac{2}{3}} = 5.78\,\mathrm{nm}$.

The orientation factor

$$\kappa^2(t) = [\cos\theta_{\mathrm{DA}}(t) - 3\cos\theta_{\mathrm{D}}(t)\cos\theta_{\mathrm{A}}(t)]^2 \tag{10}$$

depends on the three relevant angles defined in Fig. 3. The transition dipole moment orientations within the molecular frame of the dyes were chosen parallel to the ring system plane, and connecting the terminal rings of each dye (Fig. 1) [73].

Using the above framework, for all MD trajectories orientation factors $\kappa^2(t)$ and distances $R(t)$ were calculated and stored for each time step, thus obtaining time-dependent FRET rate coefficients $k_{\mathrm{T}}(t)$, which will be used below. Supplementary Video S1 shows distance, orientation factor and transfer efficiency for an exemplary trajectory.

We note that for small inter-dye distances (<2nm), when terms of higher order than the dipolar are not negligible, Eq. 8 can be replaced by multipole expansion of the coupling potential or the transition density cube method [42,74] in a straightforward manner, such that accurate FRET rate coefficients are also obtained in these cases. In the present work, the dipolar coupling potential was used.

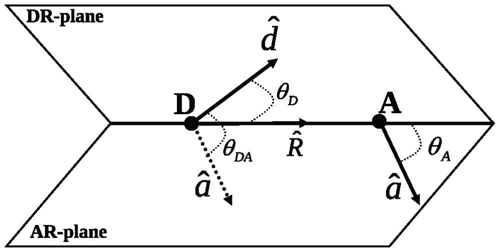


Figure 3. Geometry of dye orientations. Three angles define the orientation factor κ^2, the angle θ_{DA} between $\hat{d}$ and $\hat{a}$, and the angles θ_{D} and θ_{A} between $\hat{d}$ and $\hat{a}$, respectively, and $\hat{R}$. The DR and DA plane are defined by $\hat{R}$ and $\hat{d}$ as well as $\hat{R}$ and $\hat{a}$.
doi:10.1371/journal.pone.0019791.g003

Single Photon Generation

For direct comparison with smFRET burst counts, we developed a Monte Carlo (MC) procedure to calculate single burst FRET efficiencies from $k_{\mathrm{T}}(t)$. In the experiments, the arrival times of individual photons from single molecules were recorded. Accordingly, and following the kinetics scheme Eq. 3, multiple individual photons were generated in a Monte Carlo process (Fig. 4). For each photon, we proceeded as follows.

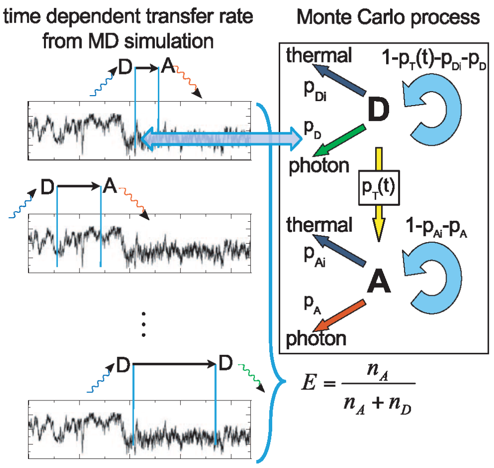


Figure 4. Photon generation by Monte Carlo. FRET transfer rate coefficient vs. time, calculated from a molecular dynamics simulation (box, left part). A random starting point on the trajectory is chosen at which the donor dye is assumed to be excited by a photon (blue). Then, for each time step the MC process on the right side is evaluated according to the corresponding probabilities until de-excitation occurs. Four de-excitation pathways are considered, thermal de-excitation of donor or acceptor (dark-blue) and respective photon emissions (donor: green; acceptor: red). The ratio of the collected donor and acceptor photons is used to calculate a FRET transfer efficiency.
doi:10.1371/journal.pone.0019791.g004

First, a random donor excitation instance was chosen from a randomly chosen trajectory (Fig. 4 left). Next, the Markov scheme in Fig. 4 (right) was iterated in time steps of Δt until either photon emission or radiationless decay occurred (see Suppl. Video S2). In the latter case, the MC run was discarded, in the former, the photon (donor or acceptor) was recorded. During each MC cycle and using an integration time step $\Delta t = 1\,\mathrm{ps}$, transitions were randomly selected according to probabilities $p_{\mathrm{Di}} = k_{\mathrm{Di}} \cdot \Delta t$ for thermal de-excitation, $p_{\mathrm{D}} = k_{\mathrm{D}} \cdot \Delta t$ for donor photon emission, $p_{\mathrm{T}} = k_{\mathrm{T}}(t) \cdot \Delta t$ for FRET transfer and $1 - p_{\mathrm{Di}} - p_{\mathrm{D}} - p_{\mathrm{T}}(t)$ for no state change. Acceptor de-excitation probabilities were calculated in the same way, but with consistent transition probabilities p_{Ai} and p_{A}, which allowed to skip the remaining Monte Carlo step and to record the emitted photon right away. All random numbers were generated with an SIMD-optimized Mersenne Twister algorithm [75,76].

In the experiment, no FRET is seen for dyes in or close to van-der-Waals contact, presumably due to quenching by electron transfer [77]. The effect of quenching at low inter-dye distances is not described with Förster theory, and therefore also not in our MC process. To correct for this, photons are rejected if the inter-dye distance is below $1\,\mathrm{nm}$ during the photon generation when comparing to experiments.

FRET Efficiency Calculation

Averaged over many MC runs, the collected de-excitation events $n_{\mathrm{Atot}} = n_{\mathrm{Ai}} + n_{\mathrm{A}}$ and $n_{\mathrm{Dtot}} = n_{\mathrm{D}} + n_{\mathrm{Di}}$ from donor and acceptor, respectively, were used to determine the average efficiency

$$E = \frac{n_{\mathrm{Atot}}}{n_{\mathrm{Atot}} + n_{\mathrm{Dtot}}}. \tag{11}$$

In experiments, only radiative de-excitation events $(n_{\mathrm{A}}, n_{\mathrm{D}})$ can be recorded. We therefore followed the same way in reconstructing the total number of de-excitation events using the respective fluorescence quantum yields,

$$n_{\mathrm{Atot}} = \frac{n_{\mathrm{A}}}{Q_{\mathrm{A}}}, \tag{12}$$

and analogously for n_{Dtot}.

To directly relate efficiency distributions from MC sampling to single molecule FRET measurements, the effect of shot noise and burst size distribution has to be taken into account properly [31,78]. Here, a sufficiently large number (>50000) of bursts has been measured, which provided sufficient statistics such that the experimental burst size distribution was used for combining the MC generated photons into bursts. After correction for quantum yield and detector efficiency, for each burst a single FRET efficiency value was calculated using Eq. 11. Collecting FRET efficiencies from many bursts yielded efficiency distributions that can be directly compared to the measured ones. As in the experiment, only bursts larger than 100 photons, after correction for the effective quantum yield, were used.

Inclusion of *cis/trans* isomer heterogeneity

So far, we have considered only one isomeric state of the proline polymer, e.g., the all-*trans* state. As has been found by NMR, however, each peptide bond undergoes isomerizations, with a small but non-negligible population in the *cis*-isomer, and with a larger *cis*-population for the terminal peptide bond at the C-terminus [45]. Because the isomerization times of minutes to hours are much longer than all other relevant time scales, we considered a weighted ensemble of all possible relevant isomerization states and performed the above MD and MC simulations with efficiency calculations separately for each isomer. Subsequently, employing $p_{\mathrm{ter}}^{\mathrm{cis}}$ and $p_{\mathrm{int}}^{\mathrm{cis}}$ from NMR experiments [45] as probabilities for the occurrence of *cis*-isomers for C-terminal and internal peptide bonds, receptively, a weighted average was obtained (Table 2).

Single-Molecule Experiments

Peptide samples were prepared as described previously [31]. Single-molecule fluorescence experiments were performed with a MicroTime 200 confocal microscope (PicoQuant, Berlin, Germany) equipped with a pulsed $485\,\mathrm{nm}$ diode laser (LDH-P-C-485B, PicoQuant) and an Olympus UplanApo 60 x/1.20 W objective. After passing through a $100\,\mathrm{mm}$ pinhole, sample fluorescence was separated by a polarizing beam splitter cube into components parallel and perpendicularly polarized with respect to the excitation light. Subsequently, both components were further divided into donor and acceptor photons by means of dichroic mirrors (585DCXR, Chroma), filtered (donor emission filters: Chroma ET525/50 M, acceptor emission filters: Chroma HQ650/100), focused on avalanche photodiodes (PerkinElmer Optoelectronics SPCM-AQR-15), and the arrival times of all detected photons were recorded using suitable counting electronics (Hydra Harp, PicoQuant, Berlin, Germany).

Results and Discussion

Time-dependent conformations of the two dyes and their mutual orientations for the three polyproline systems considered here (Fig. 2) were obtained from multiple $100\,\mathrm{ns}$ MD trajectories of the all-*trans* and *cis*-isomers. MD simulations totaling $22.1\,\mathrm{ms}$ were carried out for the different isomers, chain lengths, and temperatures (Table 1). We first focus on polyproline-20 in the all-*trans* isomer as the most stable configuration in water and analyzed two main factors relevant for the FRET efficiencies, the distance R between the two fluorophores and the orientation factor κ^2.

Distance Distributions

Fig. 5A shows the distributions of dye-to-dye distances (defined by the geometric center of the ring system) from individual $100\,\mathrm{ns}$ simulations. The fact that the distributions differ from each other shows that the individual simulations are not fully converged to represent the full all-*trans* ensemble. To improve convergence, multiple simulations were combined. The apparent differences between the individual distance distributions are mainly due to

Table 2. Isomer weights.

Isomer	Probability
all-*trans*	$p^{\mathrm{trans}} = \left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-1}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#1 cis, other trans	$p^{\mathrm{cis}_1} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#2 cis, other trans	$p^{\mathrm{cis}_2} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
$\vdots$	$\vdots$
#n-1 cis, other trans	$p^{\mathrm{cis}_{\mathrm{n}-1}} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#n cis, other trans	$p^{\mathrm{cis}_{\mathrm{n}}} = \left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-1} p_{\mathrm{ter}}^{\mathrm{cis}}$

Here, $\mathrm{cis}_1, \mathrm{cis}_2, \ldots, \mathrm{cis}_{\mathrm{n}}$ indicate the position $1, 2, \ldots, n$ of the *cis* peptide bond in the chain, starting from the amino terminus.

doi:10.1371/journal.pone.0019791.t002

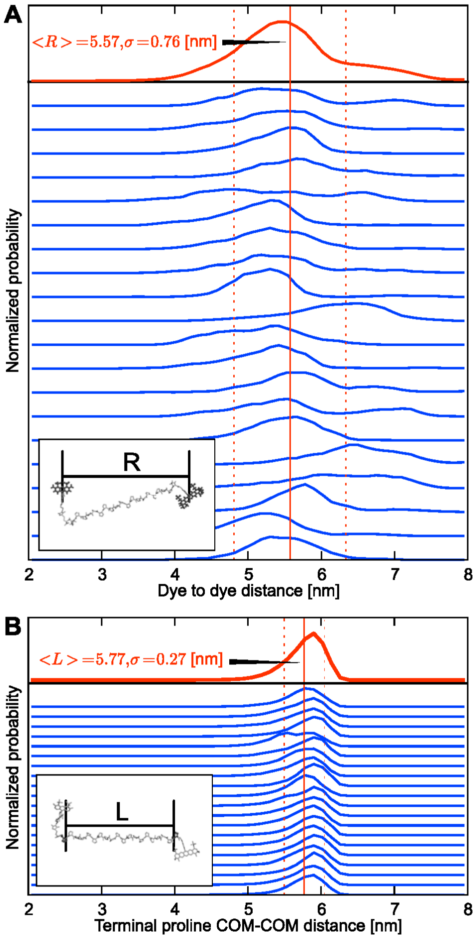


Figure 5. Distance distributions. (A) Histograms of the distances between the geometric centers of the ring systems of the two dyes for 20 all-*trans* MD simulations. (B) The distance histogram between the COM of terminal prolines from the polyproline-20 chain, for the same simulations. The insets visualize the measured distance in each plot. Respective averages are shown in red; vertical lines denote the mean and standard deviation.
doi:10.1371/journal.pone.0019791.g005

slow transitions between subpopulations of dye-conformations (for more details see section "Preferred Dye Conformations").

To better characterize the subpopulations and how they differ between the individual trajectories, the distance between two terminal proline residues center of mass (COM) was analyzed. As shown in Fig. 5B, these length fluctuations are much smaller compared to the dye-to-dye distances. In addition, the mean length of individual simulations shows only small variations.

These small length fluctuations point to considerable rigidity of the polyproline peptide, which indeed originally motivated its use as a molecular ruler. From the angular fluctuation θ of selected segment pairs, separated by length L, a persistence length $P = 18.3 \pm 0.3\,\mathrm{nm}$ was obtained via

$$P = -\frac{L}{\ln\langle\cos\theta\rangle}. \tag{13}$$

Here, 3 proline residues ($\equiv 1$ PPII helix turn) defined a segment and its tangent with a segment length of $0.93\,\mathrm{nm}$. The all-*trans* chains are indeed quite rigid and do not strongly deviate from the type II helix structure model.

Because of the stiffness of the polyproline, the observed broader distribution between the dyes mainly originate from the flexible dye linkers rather than from the flexibility of the polyproline chain.

Orientational Dye Dynamics and Orientation Factor κ^2

Figure 6 shows the κ^2 distributions derived from 20 all-*trans* simulations (gray) as well as their average (red). For comparison, an isotropic κ^2 distribution is shown (black). As shown, the individual simulations scatter considerably, with respective mean κ^2 values between $0.58 - 1.06$. Averaging over all 20 simulations, the mean κ^2 of all-*trans* simulations was 0.83 ± 0.03, and 0.80 ± 0.02 for the complete ensemble including all *cis*-isomers (Fig. 6). Both values agree within statistical error and significantly deviate from the isotropic κ^2 value of $2/3$.

As seen from the κ^2 histograms of individual MD simulations, the sampled dye geometries differ for each simulation, which underscores the importance of averaging multiple simulations. The obtained more realistic κ^2 – value shifts the effective Förster radius from $5.4\,\mathrm{nm}$ to $5.6\,\mathrm{nm}$.

Next, we determined the correlation between R and κ^2 for the 20 all-*trans* simulations and found a mean Pearson correlation coefficient of -0.13 ± 0.02. Because R and κ^2 are assumed to be uncorrelated in Försters RET theory, this finding suggests that using a distance-dependent $\langle\kappa^2\rangle(R)$ might further improve the distance reconstruction, as will be discussed below.

Table 3 shows mean auto-correlation times of different variables from the simulations (exemplary autocorrelation plot shown in

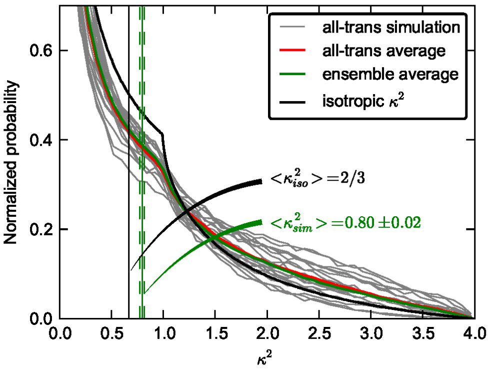


Figure 6. Distributions of the orientation factor κ^2. Each gray line shows to an orientation factor histogram from one of the 20 all-*trans* simulations at 293 K, with the average shown in red. The green curve (full ensemble) additionally includes the *cis*-isomers with appropriate weights, the green vertical line shows the corresponding average and its statistical error (dashed). The black curve shows the κ^2 histogram for an isotropic dye orientation distribution, with the well known mean value of $2/3$ (vertical black line).
doi:10.1371/journal.pone.0019791.g006

Table 3. Time scales of motions.

	mean	SEM	min	max
	[ns]	[ns]	[ns]	[ns]
R (dye-to-dye)	2.96	0.52	0.71	8.68
κ^2 (orientation factor)	0.34	0.04	0.15	0.86
L (terminal prolines)	0.80	0.20	0.30	4.12
Ω (terminal orientation)	4.96	0.86	1.06	14.55
Anisotropy decay (Alexa 488)	0.90	0.08	0.42	1.66

Autocorrelation times of all-*trans* polyproline-20 with their respective standard error of the mean (SEM), minimum and maximum. Terminal orientation Ω denotes the autocorrelation times of the cosine of the angle between the terminal proline tangent vectors.

doi:10.1371/journal.pone.0019791.t003

Suppl. Fig. S1). The orientation factor κ^2 shows the fastest decay ($0.3\,\mathrm{ns}$), whereas the terminal orientation and the dye to dye distance are in the ns regime (Table 3) and thus comparable to the donor fluorescence decay times. Calculated fluorescence anisotropy decay timescales [34,35] of $0.9\,\mathrm{ns}$ in our simulations agree with experimentally measured decay times of $0.3-0.8\,\mathrm{ns}$ [31] within the accuracy of the simulation [34] and thus indicate a correct modeling of the dye dynamics by our force field.

These autocorrelation times determine the correlation of the dye conformations and distances as probed by successive photons and, therefore also, how many structures probed by each burst are effectively statistically independent. Further, this autocorrelation time may determines the size of the sub-ensemble of conformations that is actually probed by FRET, because the fluorescence intensities of the two dyes also depend on past transfer efficiencies. We will therefore examine the influence of these effects on the quality that can be achieved for the distance reconstruction described further below.

Preferred Dye Conformations

What is the structural origin of the orientation factor κ^2 deviation from its isotropic value of $2/3$? A closer inspection of the MD simulations revealed that hydrophobic interactions of the dye linker with the proline chain enhanced the population of certain conformational sub-states, similar to previous reports[45]. This effect is more pronounced for Alexa 488 due to the longer linker. For Alexa 488, two distinct conformation sub-states (*open* and *closed*) were seen (Fig. 7).

To test the stability of these confomer ensembles, we analyzed distances and the orientation factors of the all-*trans* polyproline-20 system at elevated temperatures (303 and 313 K, Table 1). No significant impact on the values for κ^2 was found ($293\,\mathrm{K}: 0.82\pm0.03, 303\,\mathrm{K}: 0.83\pm0.05, 313\,\mathrm{K}: 0.86\pm0.04$). Also the dye-to-dye distance R showed no systematic trend towards open or closed conformations ($293\,\mathrm{K}: 5.57\pm0.08\,\mathrm{nm}, 303\,\mathrm{K}: 5.22\pm0.08\,\mathrm{nm}, 313\,\mathrm{K}: 5.37\pm0.06\,\mathrm{nm}$). For the polyproline-20 chain length L ($293\,\mathrm{K}: 5.77\pm0.01\,\mathrm{nm}, 303\,\mathrm{K}: 5.73\pm0.01\,\mathrm{nm}, 313\,\mathrm{K}: 5.70\pm0.02\,\mathrm{nm}$), a small decrease with increasing temperature was seen. In summary, the applied temperature changes neither seem to significantly influence the population ratios of the two conformations, nor the relative dye-to-dye orientations. However, due to the limited sampling, we cannot fully exclude small effects, which may arise at larger temperature changes. It will be an interesting challenge for future experimental work to directly identify the presence of the dye conformations observed here, e.g. from a broadening of fluorescence anisotropy distributions in single molecule experiments [79], or from the effect of measurements under conditions that increase the solubility of the fluorophores on the transfer efficiency histograms.

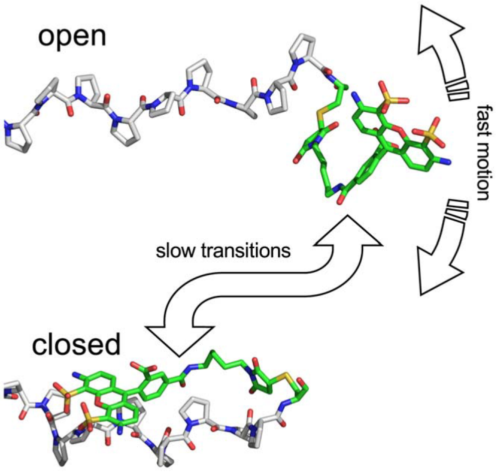


Figure 7. Conformational heterogeneity of Alexa 488. Several conformations of the Alexa 488 dye and its linker attached to the proline chain during MD simulations are seen in the simulations. For the *open* conformation, fast large amplitude motions are seen for the dye whereas hydrophobic interactions restrict the dye mobility in the *closed* conformations (one representative example is shown). Additionally slow transition between the open and closed conformations are seen.

doi:10.1371/journal.pone.0019791.g007

Efficiency Distributions from Individual Simulations

Figure 8 shows FRET efficiencies calculated separately from all 20 all-*trans* MD simulations. As already expected from the dye-to-dye distance distributions, also the mean FRET efficiencies cover a broad range from 0.27 to 0.66 with $\sigma=0.037$ to 0.043. These standard deviations σ were compared to the expected shot noise $\sigma^2=E(1-E)/n_{\mathrm{tot_L}}$ [80,81] for each simulation mean efficiency E using the lower experimental BSD limit ($n_{\mathrm{tot_L}}=100$) resulting in a width $\sigma=0.033$ to 0.036. Thus, the efficiency peak observed in the individual traces of our simulations is mainly broadened due to the photon shot noise.

Comparison of the distance distributions (Fig. 5A) with the efficiencies (Fig. 8) illustrates the effect of signal averaging over an entire fluorescence burst, subsequently referred to as ''burst averaging'. To see this, consider naive transformation from distances to efficiencies using Eq. 2, which would result in much broader efficiency distributions than those observed in Fig. 8. This narrowing is due to the combination of multiple photons, and thus also of distances, into one burst, such that each efficiency value represents a corresponding average [11]. It is this averaging, which markedly narrows obtained efficiency distributions and also obscures much of the structure seen in the distance distribution.

Isomeric Heterogeneity

To account for the isomeric heterogeneity due to the presence of *cis*-isomers, which reduce the average distance between the two

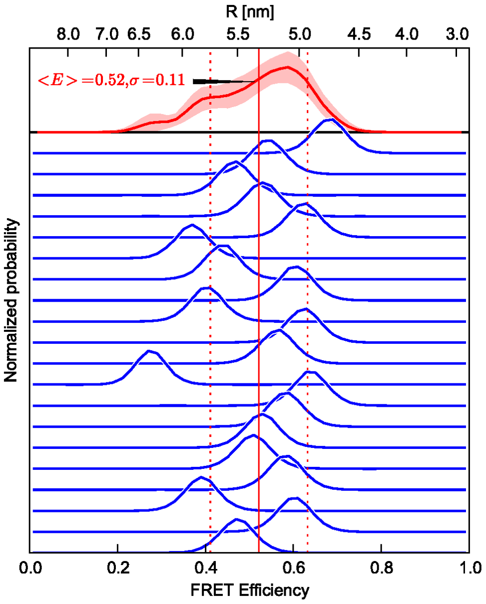


Figure 8. Spread of the efficiency amongst individual simulations. Transfer efficiency histograms (blue) obtained via MC sampling from 20 all-*trans* MD simulations of polyproline-20 at 293K. The red curve at the top depicts the efficiency combined from all 20 trajectories, where each burst is still combined from photons of one trajectory; the bootstrapping standard error, calculated from 100 random samples, is indicated by the shaded area. The vertical lines indicate the mean efficiency and its standard deviation.
doi:10.1371/journal.pone.0019791.g008

dyes [45], additional MD simulations for all possible isomers were performed (Table 1). Using the population estimate of Table 2, the full ensemble includes 5%, 8%, and 15% of isomers with more than one *cis* bond for polyprolines of length 15, 20, and 30, respectively. Thus, for proline 15 and 20, we included only the single-*cis* conformers within the ensemble. For polyproline-30, estimating the impact of multi-*cis* isomers, additionally a subset of double-*cis* isomers was considered (Table 1). In the isomer simulations, all the other bonds were kept in the *trans* configuration, and the same MD parameters and protocol as for the all-*trans* isomer were used. FRET efficiencies were then calculated as explained before.

Figure 9 shows FRET efficiency distributions and averages for the all-*trans* and *cis* polyproline-20 chains in comparison with experiment. As expected, the average efficiencies of the *cis*-chains are larger than that of the all-*trans* isomer, due to the reduced distance of the terminal prolines. The largest reduction is seen for *cis*-bonds in central positions, thus attributing measured high efficiencies to those isomers. This behavior can be captured in a simple model (Fig. 9, top), in which the *cis*-isomer is described by a kink angle α between the two stiff parts of the molecule, with distances R_A and R_D between the *cis*-bond and the respective termini, and $R = R_A + R_D$ being the all-*trans* distance between the two termini. R, was determined from the all-*trans* mean efficiency using Eq. 2 and split up on R_A and R_D for each *cis* isomer according to the *cis*-bond position. To account for the distance change due to the linker and the observed dye conformations (Fig. 7), an offset $R_D + x_0$ and $R_A - x_0$ was allowed for as an additional fit parameter. After fitting to the model to the average *cis*-efficiencies using Eq. 2, an angle of 95.5°, and and offset $x_0 = 0.27\,\mathrm{nm}$ was obtained. The resulting model is shown as green line in Fig. 9 and has to be compared to the mean efficiency values (red dots). The dashed line shows an offset of $\pm 1\,\mathrm{nm}$ in efficiency space as error estimate. The offset towards Alexa 594 $x_0 = 0.27\,\mathrm{nm}$ agrees with the deviation of the average dye-to-dye distance from the proline length $(6.01\,\mathrm{nm} - 5.57\,\mathrm{nm} = 0.44\,\mathrm{nm})$ within the accuracy of this simple model.

Next, ensemble efficiency distributions were calculated by combining *cis* and *trans* isomers according to their population in solution. Using the population of individual isomers as determined by NMR [45], $p_{\mathrm{int}}^{\mathrm{cis}}$ and $p_{\mathrm{ter}}^{\mathrm{cis}}$, weights were determined as listed in Table 2. For polyproline-20, these weights are $p^{\mathrm{trans}} = 0.6131$, $p^{\mathrm{cis}_{[1-19]}} = 0.0125, \mathrm{and}\, p^{\mathrm{cis}_{20}} = 0.0681$.

For poly-15 and polyproline-30, the same $p_{\mathrm{int}}^{\mathrm{cis}}$ and $p_{\mathrm{ter}}^{\mathrm{cis}}$ measured on polyproline-20, were applied assuming that they are not strongly influenced by the proline chain length. Because the *cis*-content is larger in polyproline-30, an error in $p_{\mathrm{int}}^{\mathrm{cis}}$ and $p_{\mathrm{ter}}^{\mathrm{cis}}$ has a larger impact on the accuracy of the ensemble composition. For example, if polyproline-30 has a $p_{\mathrm{int}}^{\mathrm{cis}}$ value of 4% instead of 2%, the multi-*cis* isomer ensemble content increases from 15% to 37%, whereas the all-*trans* isomer contribution drops from 50% to 28%. As a result, the obtained ensemble efficiency histograms sensitively depend on the value of $p_{\mathrm{int}}^{\mathrm{cis}}$ and $p_{\mathrm{ter}}^{\mathrm{cis}}$, particularly for the longer polyproline-30 chain.

As seen before, the *cis* ensemble content and thus the content of isomers with double-*cis* bonds increases with the chain length. For polyproline-30, this contribution is about 15%. To estimate the impact of double-*cis* species on the efficiency histogram, we simulated a subset of double-*cis* isomers (Table 1). The obtained weights for each chain length and isomer were used in the next step, to calculate efficiency distributions of the entire ensemble.

Combining Photons into Bursts

So far, we calculated efficiency distributions of single simulations (Fig. 8) and their accumulated histograms (Fig. 9). To calculate burst efficiencies in closer resemblance to single molecule experiments, we need to define how the recorded photons are combined, e.g. from multiple trajectories. The specific approach depends on the relative time scales of the relevant processes in the experiment and the simulation. In single molecule experiments on freely diffusing molecules, ten to hundreds of photons are recorded in each burst of several ms duration. On the simulation side, in contrast, multiple 100ns trajectories are available. We consider three different ways of combining photons into bursts and compare the resulting efficiencies to experiment.

The relevant time scales are the two autocorrelation times for the dye dynamics, namely those of the orientation factor κ^2 and of the distance R fluctuation, from hundreds of picoseconds (κ^2) to ns (R) (Table 3); further the polyproline chain dynamics of a specific isomer with the slowest motions in the 5ns range (Table 3, L and Ω), the *cis* to *trans* isomerization time ranging from minutes to hours for polyproline [82], the experimental burst recording duration of several ms and the respective inter photon times [83], as well as the simulation trajectory length of 100ns.

In the first case (burst average over fast and slow dye motions as well as the polyproline isomerization), the burst duration is assumed to be longer than all other time scales mentioned above. Accordingly, in this case, each measured burst consists of photons from the entire isomeric ensemble. To achieve a most comprehensive sampling, therefore, photons from all available trajectories

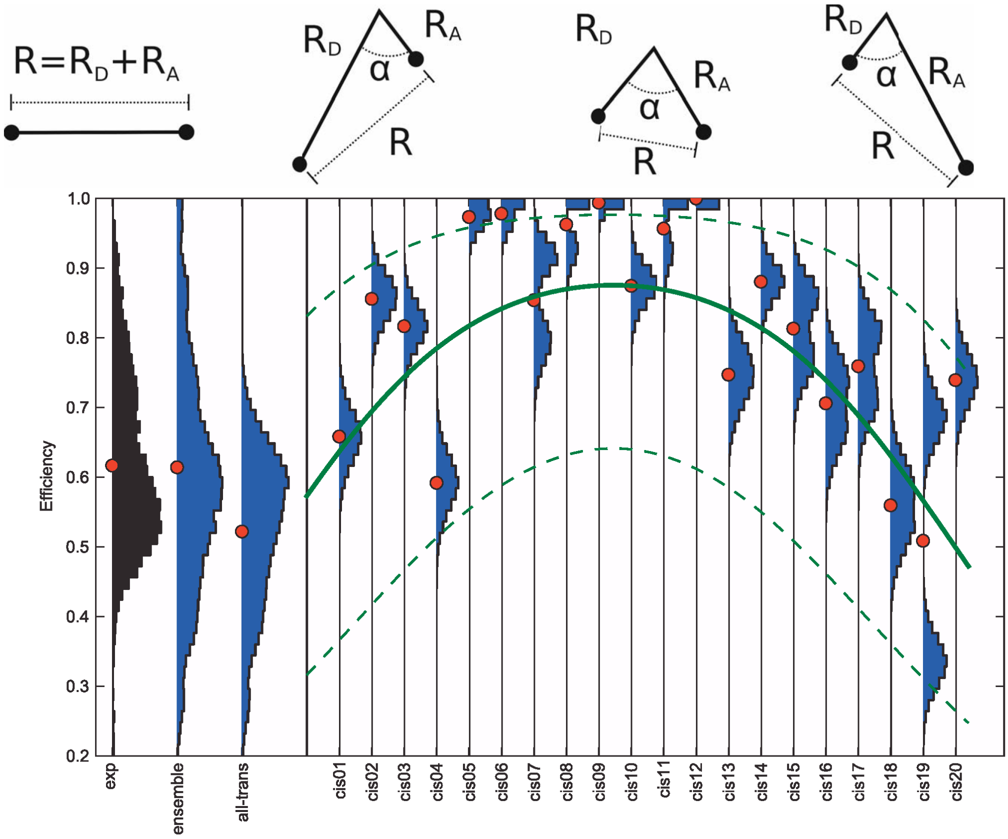


Figure 9. FRET efficiency of *trans* and *cis* isomers. Comparison between measured FRET efficiency histograms (black) and histograms computed from the simulations (blue: ensemble, all-trans and cis01 ... cis20). Red dots denote the respective mean values. The simple model sketched on top and defined in the text describes the general trend (green line) that isomers with a *cis*-bond close to the termini show lower efficiencies, whereas those with *cis*-bonds close to the polymer center tend to yield higher efficiencies. The dashed green lines estimate the spread of the average efficiencies of the *cis* simulations mirroring the spread found for the all-*trans* simulations ($\Delta R = \pm 1.0\,\mathrm{nm}$). For illustration purposes, the photons of the individual *cis* were not discarded when generated below 1 nm as described in the Methods Section. The high efficiencies observed for *cis*-6 to *cis*-12 result from dyes in contact and are quenched in the experiment.

doi:10.1371/journal.pone.0019791.g009

with their appropriate ensemble weight are combined. The blue line in Fig. 10 shows the resulting efficiency distribution as a single peak whose width is solely determined by the shot-noise. Experiments measuring ensemble efficiencies (e.g., CW in bulk) correspond to this case, except that in ensemble measurements an effectively infinite number of photons is gathered, and therefore the shot noise vanishes. For the polyproline system at hand, however, the isomerization times are long compared to the burst duration, and thus this case is not expected to apply here. Indeed, the measured efficiency distribution (Fig. 10, black) is much broader.

Accordingly, for the second case (burst average over fast and slow dye motion), we assume that the isomerization time is longer than the average burst duration, with the remaining dye and chain dynamics still being fast compared to the burst duration. In this case, all photons from a measured burst originate from one particular isomer. Because the trajectory length is much shorter than the burst duration, each burst is generated from all trajectories of a particular isomer. Figure 10, green line, shows the resulting efficiency distribution. Because in contrast to the previous case, averaging is not done over multiple isomers within each burst, as assumed above, the individual *cis* isomers contribute high efficiencies (>0.7) to the efficiency distribution (Figure 10). As shown in Fig. 10 (dashed green line), these high efficiencies are also observable in the experiment (black line). In addition, Fig. 10 reveals that the low efficiency side agrees with the experimental distribution (solid green line). However, when comparing the region around 0.7, a gap between the all-*trans* peak and the high efficiency *cis* region is present, not found in the experiment. In analogy to the comparison of this case and the above case, which averages over the polyproline isomerization, this hints at additional dynamics slower than the burst duration, averaged out in the current case.

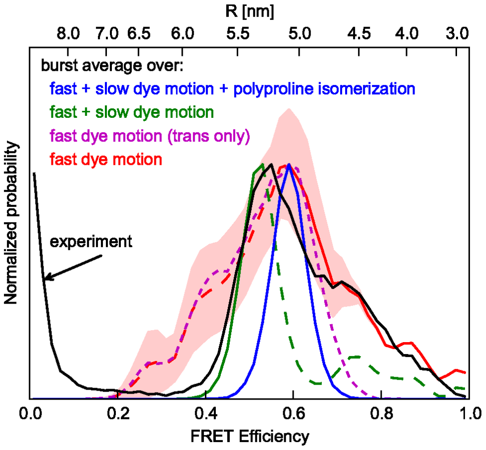


Figure 10. Combining photons into bursts. Comparison of different photon accumulation methods for a full polyproline-20 ensemble at 293 K with the experiment (black). Three different accumulation methods (colors) were considered. First, efficiencies were calculated from the full ensemble (blue), for which each photon burst has been combined from photons of all *cis* and *trans* simulations, and which therefore average over all motions and heterogeneities covered by the simulations. Second, each efficiency value was calculated from photons of all simulations of a randomly chosen isomer (green), thereby averaging over all dye motions but not over different isomers. Third, each efficiency value is derived from photons of one single trajectory, and weighted by the appropriate ensemble probability (red). The impact of the *cis*-isomers is demonstrated by comparison to the all-*trans* only efficiency histogram (magenta). The bootstrapping standard error (Fig. 8) of the all-*trans* isomers is drawn as light red area. Efficiency histograms were normalized to their maxima.
doi:10.1371/journal.pone.0019791.g010

If this is true, one would expect a better agreement for the third case considered here. In this case (burst averaging over fast dye motion only), we now assume that the dye dynamics contains additional components that are slow compared to the burst duration. An example of such a component is the transition between different conformations of the dye, e.g. the ones shown in Fig. 7. Therefore, all photons in a burst originate from a distinct dye conformation with an interconversion time larger than 100 ns. In resemblance to this, each burst is generated from one distinct simulation trajectory. The previous assumption of slow isomerization times compared to the burst recording duration is automatically contained in this case, since each trajectory contains a single distinct isomer. Figure 10 (red line) shows the resulting efficiency distribution. In contrast to the burst average over fast *and* slow dye motion, where all-*trans* and *cis*-isomers were resolvable (Fig. 10, green line), the conformational heterogeneity on time scales beyond 100 ns and thus of different simulations, is now visible as already observed in Fig. 8. As shown in Fig. 10 (red line), this heterogeneity is particularly pronounced for the all-*trans* simulations due to the largest number of simulations (Table 1) and the all-*trans* isomer being the largest fraction of the ensemble. The small numbers of simulations result in a considerable statistical error, shown as red area in Fig. 10 and calculated from the all-*trans* isomer. When comparing this result to the experiment, the high efficiency side (solid red line) with *cis*-efficiencies agrees with the experiment (black). The discrepancy (gap around 0.7) previously observed (burst average over fast *and* slow dye motion) vanishes. However, an additional low efficiency shoulder is visible not present in the experiment (dashed red line).

This deviation is not within statistical uncertainty (Fig. 10, red area) and may be due to several reasons. First, because all simulations have been started from the *open* conformation (Fig. 7), this conformation may have been oversampled. Second, although the dye dynamics described by the fluorescence anisotropy decay times agrees with the experiment, we cannot fully exclude over- or underestimation of the dye-hydrophobicity with our choice of partial charges. Third, this discrepancy can be explained by the presence of two different dye dynamics in the experiment as described below.

Overall, the low-efficiency side (<0.5 in Fig. 10) in case of burst averaging over fast *and* slow dye motions agrees well with the experiment, whereas on the high efficiency side (>0.5), better agreement is seen for burst averaging over fast dye motions only (Fig. 10, solid green and blue vs. black). From the above discussion of time scales, this finding would imply that the low efficiency side (i.e., large distances) is governed by fast dynamics, whereas parts of the slow dynamics govern the high efficiency (i.e., shorter distances) side only. Close inspection of our simulations suggests a possible structural explanation for this finding. In particular, the hydrophobic interactions between the polyproline and the Alexa 488, which give rise to the structural heterogeneity shown in Fig. 7, with very slow transitions between the *open* and *closed* conformation. In the *open* conformation, the dye-reorientation is fast compared to the burst duration and thus sampled within a single burst, in agreement with the low efficiency side (Fig. 10). In the *closed* conformations, the dye dynamics is largely restricted, with the high FRET efficiency therefore being governed by the slow transitions between these sub-states, in agreement with the observed burst averaging over fast and slow dye motions.

Next, we compare efficiency distributions for different dye-labeled proline lengths. Figure 11 shows the calculated efficiency distributions (burst averaging over fast dye motions only) from simulations with proline lengths 15, 20, and 30 (solid lines) as well as measured efficiencies for lengths of 14, 20, 27, and 33 (dashed). The general length effect, increase in efficiency for shorter prolines and vice versa, is observed.

For polyproline-15, the calculated distribution has the same narrow shape as found in the experiment, however with the simulated efficiency distribution shifted towards higher efficiencies. Purely from the length difference between polyproline-14 (experiment) and polyproline-15 (simulation), an opposite shift is expected. A similar slight discrepancy is seen for polyproline-30, where the peak should be located between the experimental peaks of polyproline-27 and -33, but is seen in Fig. 11 somewhat below polyproline-33.

While the overall agreement between simulation and experiment is good, this observed systematic deviation is striking. Apparently, compared to our simulation results, the experimental efficiencies tend to be shifted slightly towards 0.5 within both the high as well as the low efficiency regime. Overall, such behavior cannot be explained by an uncertainty in the measured R_0, which would lead to a uniform shift in one direction. With the same argument, also force field inaccuracies, which might, e.g., overestimate the hydrophobicity of the dyes and thus also the population of the *closed* conformation, are incompatible with the observed deviation. As a possible explanation one might consider a modified Förster law with, e.g., an effective power smaller than 6 in Eq. 2 (e.g. a power of ≈ 4 yields the best agreement of the simulated and experimental peak positions). Such effects have been observed previously [31] and may originate from inter-dye quenching or the breakdown of the point dipole approximation

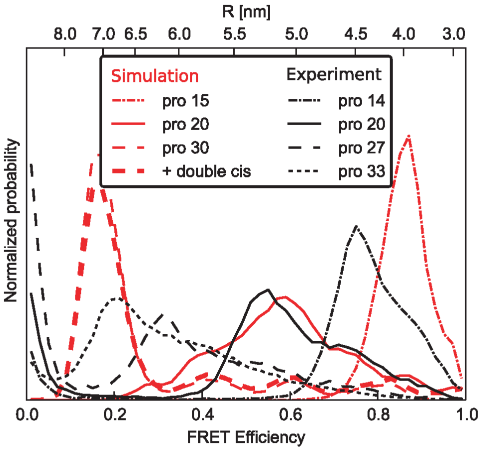


Figure 11. Comparison between proline 15, 20 and 30 and experiment. Efficiency histograms averaging over fast dye motions only (Fig. 10) are shown for three different polyproline-lengths (red, dashed-dotted, solid, dashed), and corresponding measured efficiency distributions (black). For polyproline-30, inclusion of double-*cis* isomers (see Text S1) only slightly changes the efficiency histogram (bold dashed line).
doi:10.1371/journal.pone.0019791.g011

[41,84]. As a third possible cause, decreased fluorescence lifetimes at high efficiencies leading to a stronger deviation from $\kappa^2 = 2/3$ has been discussed [31], but is already included within our simulation approach and thus unlikely to explain the deviation.

Comparing the shapes of the polyproline-30 curves, both the calculated as well as the measured efficiency distributions share shoulders reaching into the high efficiency regime. However, this shoulder is much more pronounced in the experiment than in the simulation. Closer inspection shows that the shoulder originates exclusively from *cis*-isomers. To interpret this discrepancy it is thus helpful to ask what fraction of the *cis*-population, according to the NMR results, is expected to fall into this high efficiency range. Interestingly, with the 2% *cis*-population (per bond) from NMR, and considering the fact that only about 2/3 of the cis-population contributes to the high efficiency shoulder (whereas about 1/3 contributes to the main, all-*trans* peak, see Suppl. Fig. S2), the NMR results are incompatible with the high (ca. 50%) population seen by FRET. Accordingly, a small correction of the NMR values towards higher populations of the *cis*-isomers would resolve both the discrepancy between NMR and FRET as well as that between FRET and our calculated efficiency histogram. In contrast, our neglect of multiple *cis*-conformers is unlikely to explain the discrepancy, as seen from the small effect when including the double-*cis*-species (Fig. 11, bold dashed red line) as the dominant multiple *cis*-population.

In experiments, a peak around zero efficiency is seen for all proline species. This peak originates from polyproline molecules lacking an active acceptor dye, either because of imperfect labeling or because of photobleaching of the acceptor dye during the measurement [31]. In our simulations, all molecules carry a donor and an acceptor dye, and photobleaching is not considered; the zero efficiency peak is thus absent. The clear separation of the zero efficiency peak from the rest of the signal allows us to compare only the signal from the "intact" molecules with the simulated data.

Reconstructing Distance Distributions from FRET Efficiencies

We have shown above that accurate efficiency histograms can be calculated from a combination of atomistic MD simulations and Monte Carlo photon sampling. Now we will ask the inverse question: Can the dye orientation distributions obtained from the simulations be combined with *measured* FRET efficiency histograms in such a way as to enable reconstruction of more accurate distances and, possibly, also distance distributions, than by the established $\kappa^2 = 2/3$ approximation? And if so, which accuracy can be expected at the different conceivable levels of approximations that were mentioned in the Introduction?

To address these questions, the efficiency histogram calculated from the hybrid MD/MC approach (where the distance distribution is known) as well as the single molecule FRET efficiency histogram from the experiment (where the distance distribution is unknown) were used as input for the backward calculations. The thus reconstructed distance distribution, both from the synthetic and the experimental FRET data, were then compared to the known distribution from the simulation. For each level of approximations, thereby, the impact on accuracy of the respective assumptions is quantified.

As a common framework for a proper definition of the applied approximations, we consider the most general (linear) transformation from a distance distribution $p(R)$ to an efficiency distribution $q(E)$ in terms of transfer functions $g(E,R)$,

$$q(E) = \int_0^{R_{max}} g(E,R)p(R)dR. \tag{14}$$

Each level of approximation, will be defined through an approximately specified transfer function $g(E,R)$ – or, after discretization, transfer matrix. In all cases, the all-*trans* polyproline-20 structural ensembles were used for the calculation of the transfer function as well as to generate the synthetic efficiency distribution $q(E)$; to reconstruct $p(R)$ from the experimental efficiency distribution, which involves an isomer mixture, the full structural ensemble with appropriate weights was used to calculate the transfer function (except for transfer functions g_4 and g_5). At each approximation level, $p(R)$ was then reconstructed from $q(E)$ and $g(E,R)$ by inverting a discretized version of Eq. 14.

As, generally, such inversion is numerically highly unstable, regularization assumptions are required. Here, motivated from the observation of two structural conformers (*open* and *closed* conformation, cf. Fig. 7), we assumed that $p(R)$ can be described sufficiently accurate by the sum of $n = 2$ Gaussian functions centered at R_i of width σ_i,

$$p(R) = \sum_{i=1}^{n} a_i \exp\left[-\frac{(R-R_i)^2}{2\sigma_i^2}\right]. \tag{15}$$

With this description, the χ^2 between the calculated and the reference efficiency distribution was minimized by variation of R_i and σ_i using the two array differential evolution algorithm [85]. Extension of this method to more Gaussian functions or to a more sophisticated model [86] is straightforward.

At the lowest level of refinement, the usually assumed isotropic dye orientation distribution is considered, implying $\kappa^2 = 2/3$, independent of the mutual distance between the two dyes. The efficiency distribution $q(E)$ was obtained from the donor-acceptor distance distribution $p(R)$ via the usual Förster formula, Eq. 2,

$$q(E)=p(R)\frac{dR}{dE}=-\frac{1}{6}p(R)\frac{R_0}{E^2(1/E-1)^{5/6}}. \qquad (16)$$

In the more general transfer function formalism used further below (Eq. 14), the above result (Eq. 16) is readily recovered from the transfer function

$$g_1(E,R)=\delta\left(E-\frac{1}{1+(R/R_0)^6}\right) \qquad (17)$$

shown in Fig. 12A.

Figures 13A and B show how well the respective transfer functions capture the relation between $p(R)$ and $q(E)$ as obtained from the simulations. At this first level of refinement, using the above $\kappa^2=2/3$ transfer function for both the all-trans ensemble (A) as well as for the full ensemble, containing all isomers (B), quite narrow efficiency distributions (green curves) are obtained, which are also shifted towards lower efficiencies with respect to the reference efficiency distributions (blue, black). As expected, the reconstructed distance distributions, Fig. 13E and F (same color scheme), are also shifted towards smaller distances, with the maximum being off by more than 0.5 nm. Further, the reconstructed distance distribution has a shoulder that is not seen in the reference distribution. Overall, the reconstruction is not satisfactory at this level of refinement. Figure 13C and D show, that for the all-*trans* MC and the full ensemble experimental efficiencies, respectively, adjusting of parameters in Eq. 15 led to convergence.

To quantify to which extent the assumption of an isotropic dye orientation distribution causes this discrepancy, at a second level of refinement the correct $\langle\kappa^2\rangle$ value was used, as obtained from the respective MD simulation ensemble (cf. Fig. 6). Still, this value was assumed independent of the distance between the two dyes. This approximation is described by the transfer function

$$g_2(E,R)=\delta\left(E-\frac{1}{1+(R/R_{\mathrm{adj}})^6}\right), \qquad (18)$$

with $R_{\mathrm{adj}}^6=\frac{3}{2}\langle\kappa^2\rangle R_0^6$, and $\langle\kappa^2\rangle=0.80$. As seen in Fig. 12B, this refinement results in a slight shift of the Förster curve with respect to the isotropic dye orientation approximation (Fig. 12A).

At this improved level of refinement, a slight shift of the calculated efficiency distributions towards the reference distributions is observed (red curves in Figs. 13 A, B). As a result, correspondingly improved reconstructed distance distributions are obtained (Fig. 13 E, F). However, the shapes of the efficiency curves are still too narrow, and the shoulder in the reconstructed distance distribution is still present. Apparently, these artifacts are mainly caused by further approximations not investigated so far.

Therefore, at the third level of refinement, we drop the previous assumption that the dye orientation distribution is independent of the donor-acceptor distance. Accordingly, the MD structure ensemble was split into groups according to mutual dye distance, and an average orientation factor $\langle\kappa^2\rangle_R$ was calculated separately for every group, i.e., as a function of R. Note that this distance dependent orientation factor

$$\langle\kappa^2\rangle(R)=\langle(\cos\theta_{\mathrm{DA}}-3\cos\theta_{\mathrm{D}}\cos\theta_{\mathrm{A}})^2\rangle(R) \qquad (19)$$

differs from the previous ones in that it captures correlations between the dye orientation distribution and the donor-acceptor distances. This can be used to construct the transfer function

$$g_3(E,R)=\delta\left(E-\frac{1}{1+(R/R_{\mathrm{red}})^6\cdot(1/\langle\kappa^2\rangle(R))}\right), \qquad (20)$$

defining $R_{\mathrm{red}}^6=R_0^6\cdot 3/2$.

As seen in Fig. 12C, the resulting transfer function is not strictly monotonic any more, such that the inverse transformation to $R(E)$ is not straightforward and, the above regularization techniques need to be applied.

This refinement step yields a marked improvement of both peak position and shape of the obtained efficiency distributions (Figs. 13 A and B, cyan). Only a slight peak shift towards lower efficiencies remains for the all-*trans* ensemble (Fig. 13 A), as is also seen for the experimental efficiencies in Fig. 13B. Also for the distance reconstruction, the dominant peak is now at the correct position in both cases (Fig. 13 E, F), although the second peak in the synthetic distance reconstruction using the all-*trans* ensemble still remains and leads to an overestimate of the distribution for smaller distances.

So far, our transfer functions uniquely defined the efficiency E for each distance R. Before continuing with further refinement steps, we demonstrate how the experimental shot noise impacts the reconstruction of distances. Two fundamental approaches have been used so far to calculate the shot noise contribution via numeric solution [78,80] or via simulation [87,88]. Mathematically, the shot noise free efficiency distribution $\tilde{q}(E)$ is convoluted with a shot noise kernel S resulting in an efficiency distribution $q(E)$ including the shot noise. This convolution

$$q(E)=\int\tilde{q}(E')S(E-E')dE', \qquad (21)$$

broadens the underlying efficiency distribution $\tilde{q}(E)$ to $q(E)$. Because of this broadening, the reconstructed distance distribution $p(R)$ is narrowed when shot noise is taken into account. In analogy to image reconstruction from a de-focused image by inversion of the convolution with the appropriate image transfer function, the achieved accuracy and the ability to recover finer details of the original distance distribution are limited by the information loss due to convolution of the shot noise kernel S with the transfer function $g(E,R)$, Eq. 14.

Since determining the shot noise kernel S of an experimental BSD is non-trivial, the experimental shot noise (bursts ≥ 100) was included in the transfer function as follows. Each distance bin of the transfer matrix (columns in Fig. 12) was randomly sampled by 1200 bursts from the experimental BSD. The target efficiency for each burst was directly calculated from the transfer function (g_1, g_2, g_3) or randomly picked from the efficiency distribution (for the following refinement steps). According to the target efficiency, donor and acceptor photons were randomly generated, and the obtained burst efficiency was then recorded in the transfer function. Figure 12D and H illustrate the impact of an experimental shot noise (bursts ≥ 20) on transfer functions (Fig. 12C and G). Comparison of C and D illustrates, that after the inclusion of shot noise, the transfer function not uniquely defines an efficiency E value for each distance R, but instead an efficiency distribution. The here observed effect of the BSD on the transfer function is purely of stochastic origin, whereas a similar but independent effect will be seen in the following refinement level.

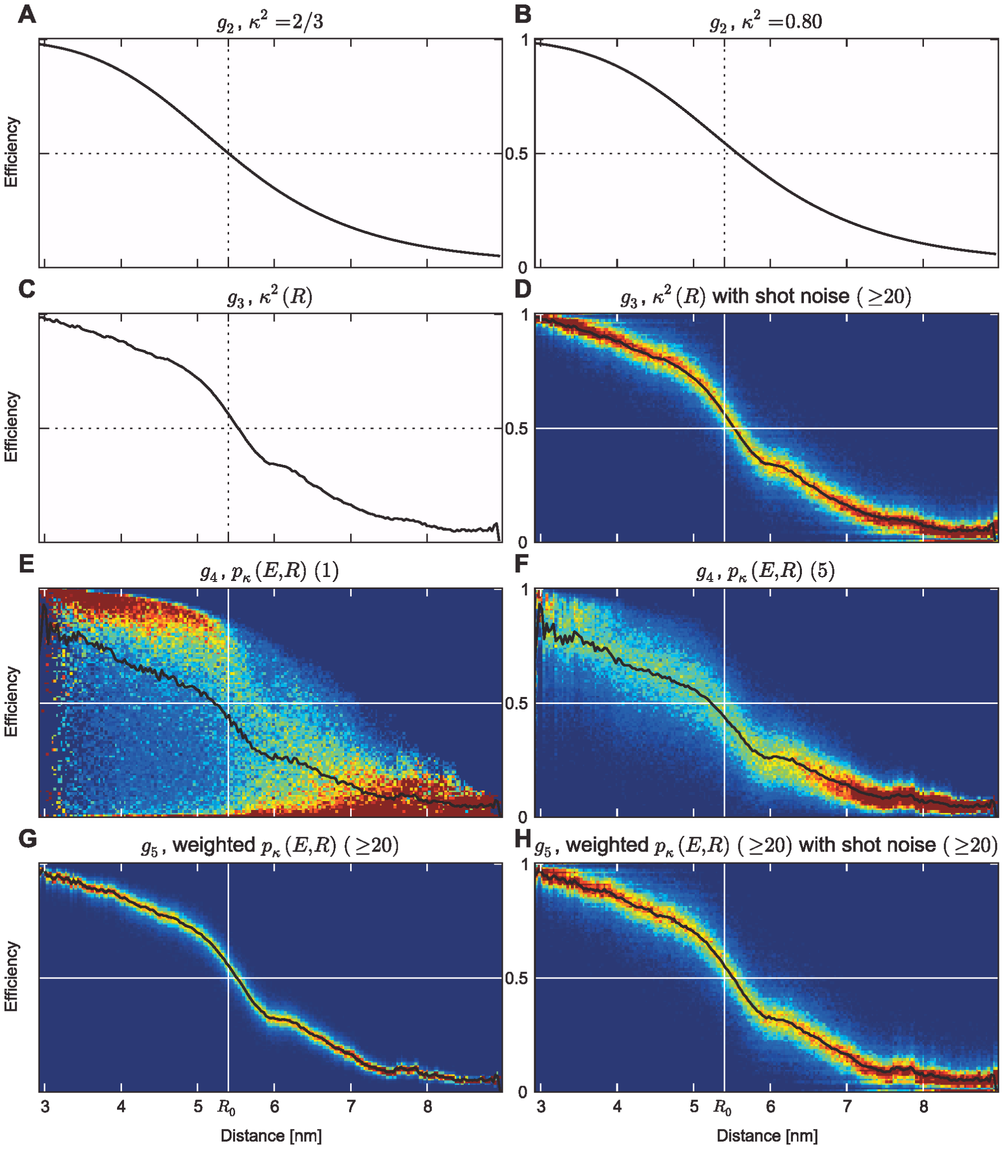


Figure 12. Transfer functions g_i at increasingly refined approximation levels. Transfer functions A, B, and C are shown as black curves; the remaining transfer functions are shown color-coded, with averages highlighted as black curves. Transfer function g_1 was calculated using the assumption of $\kappa^2 = 2/3$ (A). For g_2, κ^2 was adjusted to represent the ensemble average in the simulations (B). g_3 includes the distance dependency of κ^2 without (C) and with (D) shot noise derived from the experimental BSD (burst size or lower burst size cutoff given in brackets). In contrast to a distance dependent averaged κ^2, g_4 includes the κ^2 distributions at each distance without (E) and including averaging within a burst (F). In g_5, the time dependent photon emission (Fig. 14) is included, shown without (G) and with experimental shot noise (H).
doi:10.1371/journal.pone.0019791.g012

To motivate this level of refinement, recall that in all levels of refinement considered so far the full structure ensemble has been used to calculate appropriate averages for the orientation factor κ^2. This approach implies the salient assumption that each single burst samples the same dye orientation distribution – which, however, holds true only if all components of the dye motion are much faster

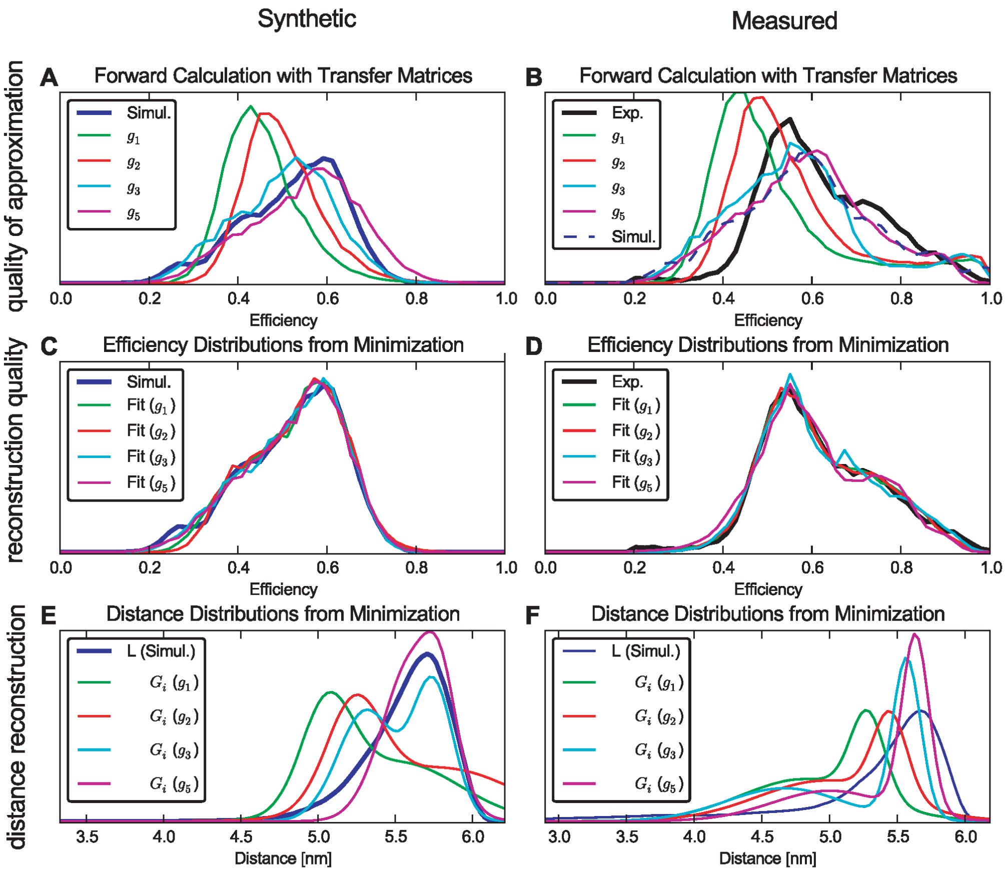


Figure 13. Distance reconstruction from efficiencies. Reconstruction of distance distributions from synthetic efficiencies with known distance distributions from simulations, and measured efficiencies with unknown distance distribution. Two ensembles were considered: The left column consists of all-*trans* polyproline-20 at 293 K. The right column also includes the *cis*-isomers with appropriate weights and uses experimental efficiencies as reference. In the first row (A and B), the efficiency distributions obtained from multiplying the transfer matrices (discretized transfer function) with the distance distributions obtained from simulations are shown together with efficiencies derived from our simulations and experiment as reference. The second row (C and D) depicts the efficiency obtained by optimizing the parameters of two Gaussian distance distributions as a measure for the reconstruction quality. The efficiencies were calculated by multiplying the transfer matrix with the Gaussian distance distributions with optimized parameters (see Text S1). The distance distributions obtained from reconstruction and simulation are shown in the third row (E and F). In all graphs, the reference is plotted with a bold line. Notably the experimental reference distance is inaccessible in F. The employed transfer matrices include experimental shot noise.
doi:10.1371/journal.pone.0019791.g013

than the burst duration. As has been shown from the above comparison between measured efficiency distributions and those obtained from three different structure ensembles, there are slow components of the dye motion, which may render this salient assumption questionable. Our last level of refinement attempts to include the dominant effect of this limited dye orientation sampling within the transfer function. Note, however, that a rigorous treatment of this effect would require to go beyond the limits of the transfer function framework, and here is only achieved conceptually by our explicit hybrid MD/MC simulation approach.

At this refinement level, accordingly, the columns of the transfer matrix are formed from distance dependent transfer efficiency distributions $p_\kappa(\kappa^2,R)$ rather than single valued R-dependent averages $\langle\kappa^2\rangle(R)$, from which the transfer function $g_4(E,R)$ is derived as

$$g_4(E,R)=\frac{p_\kappa(\kappa^2,R)}{\int d\kappa^2 p_\kappa(\kappa^2,R)}\cdot\frac{d\kappa}{dE}. \qquad (22)$$

Here, the integral over κ^2 in the denominator of g_4 normalizes the probability distribution on the distances and $d\kappa/dE$ transforms $p_\kappa(\kappa^2,R)$ to $p_\kappa(E,R)$. The normalized transfer function is obtained from orientation factor histograms for different distances from the MD trajectory ensemble, applying Eq. 2.

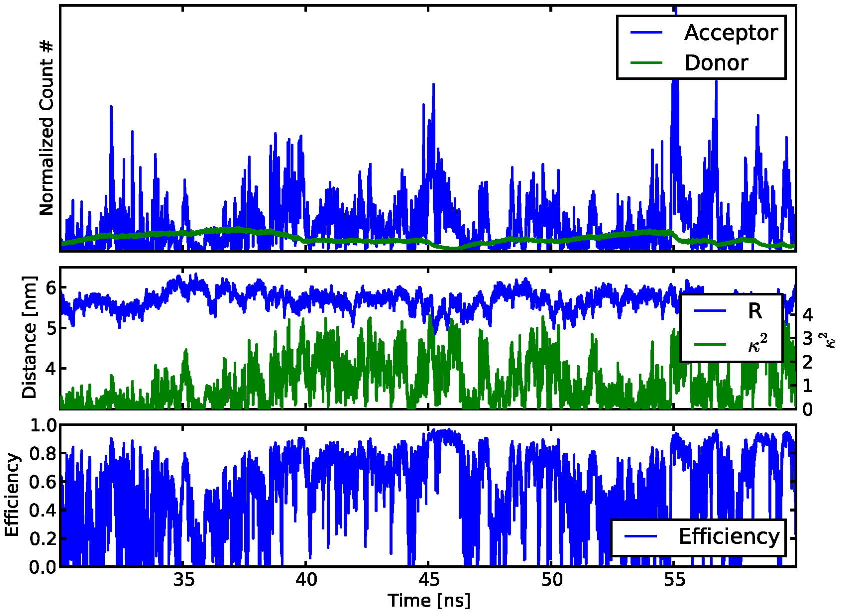


Figure 14. Time dependent photon emission along a single trajectory. Top: normalized acceptor (blue) and donor (green) photon count for time independent excitation probability. Mid: corresponding distance $R(t)$ and orientation factor $\kappa^2(t)$. Bottom: resulting time dependent instantaneous efficiency.

doi:10.1371/journal.pone.0019791.g014

Figure 12E shows the transfer function resulting from the samples of our simulations. Notably, there is a distance dependent maximum efficiency due to the κ^2 range from 0 to 4. The samples of $p_\kappa(\kappa^2,R)$ from our simulations each determine the efficiency samples of $p_\kappa(E,R)$. In experiments, however, efficiencies are determined using multiple photons. Thus, the efficiencies in the transfer function need to be averaged over multiple (κ^2,R) samples according to the BSD. Figure 12F shows this effect for a constant burst size of 5. As seen, this burst size dependent averaging introduces a narrowing of the transfer function, independent of the photon shot noise.

Figure 14 motivates and illustrates the next level of refinement. Shown are the donor and acceptor photon counts from a trajectory, extensively sampled with photons created in our MC process. In high efficiency regions with efficiency values in the lower plot close to one (e.g., around 45 and 55 ns), a marked depletion of donor and acceptor photon counts (dips in the green curve in the upper plot) is observed. As a result, also the mean intensity I_{avg} depends on κ^2 and R and thus affects the probability of obtaining a photon from a distinct (κ^2,R) conformation. Thus, for each instant t, the orientation $(\kappa^2(t),R(t))$ depends on the history of orientations within the fluorescence lifetime. Because in the construction of g_4, the probability distribution of dye orientations $p_\kappa(\kappa^2,R)$ is only normalized at each distance, this memory is not described. By applying appropriate intensity weights, our last, most realistic transfer function g_5 includes also this effect.

$$g_5(E,R)=\frac{p_\kappa(\kappa^2,R)\cdot I_{\text{avg}}(\kappa^2,R)}{\int d\kappa^2 p_\kappa(\kappa^2,R)\cdot I_{\text{avg}}(\kappa^2,R)}\cdot\frac{d\kappa}{dE}. \tag{23}$$

In our transfer function construction, the intensity of each (κ^2,R) sample was determined by extensive photon sampling of our trajectories. Thereby, the adsorption events were equally distributed over the whole trajectory and the emission times of the photons was recorded. The (κ^2,R) samples were then weighted according to their total emitted photon count. Notably, the samples are implicitly weighted according to their efficiency history in experiments. In Fig. 12G, a shift towards higher efficiencies as an effect of this weighting is seen. To reduce computationally expensive photon sampling of the trajectories, g_4 and g_5 were calculated from 20 all-*trans* simulations only.

Applying this transfer function g_5 to our known distance distributions to asses the quality of approximations results in efficiency distributions only slightly different from the ones for g_3 (Fig. 13 A, B). Nevertheless, as seen in Fig. 13, the high efficiencies in the experimental ensemble were reproduced better than for g_3. When using g_5 for the reconstruction of distances using the synthetic efficiencies, the best agreement with the distance distribution from the simulations was found (Fig. 13 E). Also, the reconstructed peak location using the experimental efficiencies is slightly closer to the peak from the simulations (Fig. 13 F).

Overall, in the experimental reconstruction, all distance distributions of different refinement levels are shifted towards lower distances in comparison to the simulation distance distribution. This agrees with the observation of low efficiency overestimation shown in Fig. 10.

These tests demonstrate, that a markedly improved reconstruction over the established approaches is achieved by inclusion of dye motion and photon statistics obtained by our hybrid simulation approach of simulated data. Further, by using step-by-step refined approximation levels for the transfer functions, a systematic improvement of the inverse distance reconstruction is achieved for the polyproline system.

Conclusions

We have demonstrated that structural information on the dynamics of FRET dye pairs from MD simulations improves the reconstruction of distances and distance distributions from experimental FRET efficiency distributions over the usual $\kappa^2 = 2/3$ approximation, which assumes isotropic and uncorrelated distributions of the dye transition dipole orientations. A hybrid MC/MD method was developed and tested, which uses this structural information in combination with a Monte Carlo description of photon absorption, FRET-transfer, and emission, to calculate quantitative efficiency distributions. Based on the obtained good agreement with measured efficiency distributions of polyproline constructs, we have investigated several levels of approximation, resting on the particular relation of the different relevant time scales of the experiment and of the simulations. For the system at hand, this analysis revealed a previously unknown slow component of the dye movement. Our analysis further highlights that careful consideration of the time scales of the involved processes is crucial, and offers a framework that is flexible enough to capture the different time scale relationships expected for a broad range of systems. Unexpectedly, already for the simple polyproline system at hand, where the dyes are usually assumed to be sufficiently flexible to justify the established $\kappa^2 = 2/3$ approximation, severe deviations were seen. Our results suggest that for FRET dye pairs attached to proteins or DNA/RNA complexes, the orientational dynamics are typically more restricted due to sterical hindrance and electrostatic interactions, a simulation approach like the one developed here is essential.

Supporting Information

Figure S1 Autocorrelation decay times of multiple parameters. R is the inter-dye distance, L the chain end-to-end distance and κ^2 the orientation factor. The 3D, 2D and 2nd Legendre Polynomial of 2D (Anisotropy decay) was determined from Alexa 488. The decay here is from an all-*trans* polyproline-20 simulation.
(EPS)

Figure S2 Polyproline-30 *cis*-isomer efficiencies. For each isomer, the normalized probability is shown.
(EPS)

Video S1 Distance and Orientation factor from simulations. For illustration, a fragment of 10 ns simulation time from polyproline-15 with the two dyes attached is shown as example for the dynamics. The box in the bottom shows the time dependent orientation and distance as well as the resulting FRET efficiency. The position in the time trace is shown as moving red bar in the box.
(M4V)

Video S2 FRET from Monte Carlo and simulation trajectories. In addition to Video S1, exemplary excitation and de-excitation events are shown. Both competing pathways, de-excitation of donor, as well as the alternate pathway via FRET followed by acceptor de-excitation are displayed.
(M4V)

Text S1
(PDF)

Table S1
(PDF)

Acknowledgments

We thank Ulrike Gerischer, Jan Neumann, Jan Opfer, Stephanus Fengler, Stanislav Kalinin and Christian Blau for carefully reading the manuscript and Suren Felekyan, Brent P. Krueger as well as Daniel Nettels for helpful discussions.

Author Contributions

Conceived and designed the experiments: MH NL BS CAMS HG. Performed the experiments: MH NL BS HG DH. Analyzed the data: MH NL BS. Contributed reagents/materials/analysis tools: MH NL BS CAMS HG DH. Wrote the paper: MH NL BS CAMS HG.

References

1. Förster T (1948) Zwischenmolekulare Energiewanderung und Fluoreszens. Annalen der Physik 2: 55–75.
2. Stryer L, Haughland RP (1967) Energy transfer - a spectroscopic ruler. Proc Natl Acad Sci U S A 58: 719–726.
3. Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CAM (2010) Accurate single-molecule FRET studies using multiparameter fluorescence detection. Methods Enzymol 475: 455–514.
4. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, et al. (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proc Natl Acad Sci U S A 93: 6264–6268.
5. Shera EB, Seitzinger NK, Davis LM, Keller RA, Soper SA (1990) Detection of single fluorescent molecules. Chem Phys Lett 174: 553 - 557.
6. Berglund AJ, Doherty AC, Mabuchi H (2002) Photon statistics and dynamics of fluorescence resonance energy transfer. Phys Rev Lett 89: 068101.
7. Meer BWVD, III GC, Chen SYS (1994) Resonance Energy Transfer: Theory and Data. John Wiley & Sons, 1 edition.
8. Kalinin S, Felekyan S, Valeri A, Seidel CAM (2008) Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. J Phys Chem B 112: 8361–8374.
9. Neubauer H, Gaiko N, Berger S, Schaffer J, Eggeling C, et al. (2007) Orientational and dynamical heterogeneity of rhodamine 6G terminally attached to a DNA helix revealed by NMR and singlemolecule fluorescence spectroscopy. J Am Chem Soc 129: 12746–12755.
10. Wozniak A, Nottrott S, Kuhn-Holsken E, Schröder G, Grubmüller H, et al. (2005) Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. RNA-A Publication of the RNA Society 11: 1545–1554.
11. Sindbert S, Kalinin S, Nguyen H, Kienzler A, Clima L, et al. (2011) Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity. J Am Chem Soc 133: 2463–2480.
12. Wozniak AK, Schröder GF, Grubmüller H, Seidel CAM, Oesterhelt F (2008) Single-molecule FRET measures bends and kinks in DNA. Proc Natl Acad Sci U S A 105: 18337–18342.
13. Brunger AT, Strop P, Vrljic M, Chu S, Weninger KR (2010) Three-dimensional molecular modeling with single molecule FRET. J Struct Biol 173: 497–505.
14. Düser MG, Bi Y, Zarrabi N, Dunn SD, Börsch M (2008) The proton-translocating a subunit of F0F1-ATP synthase is allocated asymmetrically to the peripheral stalk. J Biol Chem 283: 33602–33610.
15. Schröder G, Grubmüller H (2004) FRETsg: biomolecular structure model building from multiple FRET experiments. Comput Phys Commun 158: 150–157.
16. Uphoff S, Holden SJ, Reste LL, Periz J, van de Linde S, et al. (2010) Monitoring multiple distances within a single molecule using switchable FRET. Nat Methods 7: 831–836.
17. Ha T (2001) Single-molecule fluorescence resonance energy transfer. Methods 25: 78–86.
18. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. Curr Opin Struct Biol 18: 16–26.
19. Widengren J, Kudryavtsev V, Antonik M, Berger S, Gerken M, et al. (2006) Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. Anal Chem 78: 2039–2050.
20. Gopich IV, Szabo A (2003) Single-macromolecule fluorescence resonance energy transfer and freeenergy profiles. J Phys Chem B 107: 5058–5063.
21. Margittai M, Widengren J, Schweinberger E, Schröder G, Felekyan S, et al. (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. Proc Natl Acad Sci U S A 100: 15516–15521.
22. Hofmann H, Hillger F, Pfeil SH, Hoffmann A, Streich D, et al. (2010) Single-molecule spectroscopy of protein folding in a chaperonin cage. Proc Natl Acad Sci U S A 107: 11793–11798.

23. Schröder GF, Grubmüller H (2003) Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments. J Chem Phys 119: 9920–9924.
24. Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing single-molecule FRET dynamics with probability distribution analysis. Chemphyschem 11: 2209–2219.
25. Hillger F, Hänni D, Nettels D, Geister S, Grandin M, et al. (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. Angew Chem Int Ed Engl 47: 6184–6188.
26. Van Oort B, Murali S, Wientjes E, Koehorst RBM, Spruijt RB, et al. (2009) Ultrafast resonance energy transfer from a site-specifically attached fluorescent chromophore reveals the folding of the N-terminal domain of CP29. Chem Phys 357: 113–119.
27. Yefimova SL, Gural'chuk GY, Sorokin AV, Malyukin YV, Borovoy IA, et al. (2008) Hydrophobicity effect on interactions between organic molecules in nanocages of surfactant micelle. J Appl Spectrosc 75: 658–663.
28. Sanborn ME, Connolly BK, Gurunathan K, Levitus M (2007) Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. J Phys Chem B 111: 11064–11074.
29. Ranjit S, Gurunathan K, Levitus M (2009) Photophysics of backbone fluorescent DNA modifications: reducing uncertainties in FRET. J Phys Chem B 113: 7861–7866.
30. Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. J Phys Chem B 114: 6197–6206.
31. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the spectroscopic ruler revisited with single-molecule fluorescence. Proc Natl Acad Sci U S A 102: 2754–2759.
32. Lakowicz CDGJR, ed (2006) Reviews in Fluorescence / Annual volumes 2006 (Reviews in Fluorescence). Springer, 1 edition.
33. Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, et al. (2003) Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. Proc Natl Acad Sci U S A 100: 1655–1660.
34. Schröder GF, Alexiev U, Grubmüller H (2005) Simulation of fluorescence anisotropy experiments: Probing protein dynamics. Biophys J 89: 3757 - 3770.
35. Henry ER, Hochstrasser RM (1987) Molecular dynamics simulations of fluorescence polarization of tryptophans in myoglobin. Proc Natl Acad Sci U S A 84: 6142–6146.
36. Rindermann JJ, Akhtman Y, Richardson J, Brown T, Lagoudakis PG (2011) Gauging the flexibility of fluorescent markers for the interpretation of fluorescence resonance energy transfer. J Am Chem Soc 13: 1857–1871.
37. Corry B, Jayatilaka D, Martinac B, Rigby P (2006) Determination of the orientational distribution and orientation factor for transfer between membrane-bound fluorophores using a confocal microscope. Biophys J 91: 1032–1045.
38. VanBeek DB, Zwier MC, Shorb JM, Krueger BP (2007) Fretting about FRET: correlation between kappa and R. Biophys J 92: 4168–78.
39. Di Fiori N, Meller A (2010) The effect of dye-dye interactions on the spatial resolution of singlemolecule FRET measurements in nucleic acids. Biophys J 98: 2265–2272.
40. Corry B, Jayatilaka D (2008) Simulation of structure, orientation, and energy transfer between AlexaFluor molecules attached to MscL. Biophys J 95: 2711–2721.
41. Dolghih E, Ortiz W, Kim S, Krueger BP, Krause JL, et al. (2009) Theoretical studies of short polyproline systems: Recalibration of a molecular ruler. J Phys Chem A 113: 4639–4646.
42. Speelman AL, Munosz-Losa A, Hinkle KL, VanBeek DB, Mennucci B, et al. (2011) Using molecular dynamics and quantum mechanics calculations to model fluorescence observables. J Phys Chem A in press.
43. Unruh JR, Kuczera K, Johnson CK (2009) Conformational heterogeneity of a leucine enkephalin analogue in aqueous solution and sodium dodecyl sulfate micelles: comparison of time-resolved FRET and molecular dynamics simulations. J Phys Chem B 113: 14381–14392.
44. Schuetz P, Wuttke R, Schuler B, Caisch A (2010) Free energy surfaces from single-distance information. J Phys Chem B 114: 15227–15235.
45. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A, et al. (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. Proc Natl Acad Sci U S A 104: 18964–18969.
46. Berger A, Kurtz J, Katchalski E (1954) Poly-L-Proline. J Am Chem Soc 76: 5552–5554.
47. Harrington WF, Sela M (1958) Studies on the structure of poly-L-proline in solution. Biochim Biophys Acta 27: 24–41.
48. Dorman DE, Torchia DA, Bovey FA (1973) Carbon-13 and proton nuclear magnetic resonance observations of the conformation of poly(L-proline) in aqueous salt solutions. Macromolecules 6: 80–82.
49. Zagrovic B, Lipfert J, Sorin EJ, Millett IS, van Gunsteren WF, et al. (2005) Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci U S A 102: 11698–11703.
50. Doose S, Neuweiler H, Barsch H, Sauer M (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. Proc Natl Acad Sci U S A 104: 17400–17405.
51. Hoffmann A, Nettels D, Clark J, Borgia A, Radford S, et al. Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: Recurrence analysis of single particles (RASP). Phys Chem Chem Phys, (in press).
52. Panchuk-Voloshina N, Haugland RP, Bishop~nStewart J, Bhalgat MK, Millard PJ, et al. (1999) Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. J Histochem Cytochem 47: 1179–1188.
53. Cownan PM, McGavin S (1955) Structure of poly-L-proline. Nature 176: 501–503.
54. Moradi M, Babin V, Roland C, Darden TA, Sagui C (2009) Conformations and free energy landscapes of polyproline peptides. Proc Natl Acad Sci U S A 106: 20746–20751.
55. Grathwohl C, Wüthrich K (1981) NMR studies of the rates of proline cis-trans isomerization in oligopeptides. Biopolymers 20: 2623–2633.
56. Jorgensen WL, Madura JD (1985) Temperature and size dependence of Monte Carlo simulations of TIP4P water. Mol Phys 56: 1381–1392.
57. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225–11236.
58. Carlson HA, Nguyen TB, Orozco M, Jorgensen WL (1993) Accuracy of free-energies of hydration for organic molecules from 6-31G-asterisk-derived partial charges. J Comput Chem 14: 1240–1249.
59. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. J Comput Chem 5: 129–145.
60. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 03, B.04. Gaussian, Inc.,Wallingford, CT, 2004.
61. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS - a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 91: 43–56.
62. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4: 435–447.
63. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26: 1701–1718.
64. Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20: 786–798.
65. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690.
66. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126: 014101.
67. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52: 7182–7190.
68. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18: 1463–1472.
69. Darden T, York D, Pedersen L (1993) Particle mesh ewald - an n.log(N) method for ewald sums in large systems. J Chem Phys 98: 10089–10092.
70. Schuler B (2007) Application of single molecule förster resonance energy transfer to protein folding. Methods Mol Biol 350: 115–138.
71. Braslavsky SE, Fron E, Rodrguez HB, Romn ES, Scholes GD, et al. (2008) Pitfalls and limitations in the practical use of förster's theory of resonance energy transfer. Photochem Photobiol Sci 7: 1444–1448.
72. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419: 743–747.
73. Rusinova E, Tretyachenko-Ladokhina V, Vele OE, Senear DF, Ross JBA (2002) Alexa and Oregon green dyes as fluorescence anisotropy probes for measuring protein-protein and protein-nucleic acid interactions. Anal Biochem 308: 18–25.
74. Krueger BP, Scholes GD, Fleming GR (1998) Calculation of couplings and energy-transfer pathways between the pigments of LH2 by the ab initio transition density cube method. J Phys Chem B 102: 9603–9604.
75. Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 8: 3–30.
76. Saito M, Matsumoto M (2008) SIMD-oriented fast mersenne twister: a 128-bit pseudorandom number generator. In: Keller A, Heinrich S, Niederreiter H, eds (2008) Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer Berlin Heidelberg. pp 607–622.
77. Brune R, Doose S, Sauer M (2007) Analyzing the influence of contact-induced quenching processes on Förster resonance energy transfer (Proceedings Paper). In: Popp J, von Bally G, eds. Biophotonics 2007: Optics in Life Science volume 6633 of Proceedings of SPIE;doi:10.1117/12.727853.
78. Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. J Phys Chem B 110: 6970–6978.
79. Kalinin S, Felekyan S, Antonik M, Seidel CAM (2007) Probability distribution analysis of singlemolecule fluorescence anisotropy and resonance energy transfer. J Phys Chem B 111: 10253–10262.
80. Gopich I, Szabo A (2005) Theory of photon statistics in single-molecule forster resonance energy transfer. J Chem Phys 122: 14707.
81. Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. J Phys Chem B 111: 12925–12932.

Supplementary Materials

Cis-model

To calculate structural properties of the *cis*-simulations, the obtained average efficiencies were fitted to a simple geometrical model.

$$R_{\text{tot}} = \left(\frac{1}{E_{\text{tot}}} - 1\right)^{\frac{1}{6}} \cdot R_0 \tag{24}$$

$$R_A(i, x_0) = i \cdot \frac{R_{\text{tot}}}{20} - \frac{R_{\text{tot}}}{40} - x_0 \, [\forall i \in \{1, \ldots, 20\}] \tag{25}$$

$$R_D(i, x_0) = R_{\text{tot}} - R_A(i, x_0) \tag{26}$$

$$R_x(i, x_0, \alpha) = \sqrt{R_D^2 + R_A^2 + 2 \cdot R_D \cdot R_A \cos\alpha} \tag{27}$$

Its basis is a triangle, containing donor and acceptor arm lengths R_D and R_A (Eq. 25 and 26) from the *cis*-kink at the i-th position and the angle between the two arms α. The third side in the triangle is the end-to-end distance as dye separation R_x (Eq. 27). Additionally, an asymmetry x_0 is added to one and subtracted from the other arm length. The all-*trans* average efficiency is used to define the length of our chain R_{tot} according to Eq. 1 in our model. The position of the kink, and thus the length of the two arms, is determined by splitting R_{tot} into 20 segments and choosing the midpoint as kink. During the fit, the parameters α and x_0 are optimized.

Simulation box creation

First, the polyproline was aligned along the principal component of the terminal prolines (x-axis) and embedded in a box with a minimal distance of 1.5 nm to the boundaries. Then, the box vectors perpendicular to the principal component (y and z-direction) were symmetrized by applying either the larger one or a length of at least 6.2 nm to the y and z box vector length. This allows application of weak restraints to prevent rotation. Finally, the box vector in x-direction was enlarged by 1.8 nm to suppress interaction of the dyes with the periodic image of their counterpart (Fig. 2B).

Histograms

For efficiency histograms 50 histogram bins were used, evenly spaced from 0 to 1. Additionally, a random number in the range ±0.00001 was added to all efficiency values. This prevents problems at the bin boundaries e.g. at 0.5, where all efficiencies with equal donor and acceptor ratio are put in bin # 25. In experimental setups, the raw data exhibits the same problems, however it is usually masked by the instrument corrections.

Persistence length calculation

Persistence length was determined from segment tangents to the chain. These tangents were defined by the vector $\mathrm{C} - \alpha_i$ and $\mathrm{C} - \alpha_i + (3 \cdot n)$ due to the periodicity of 3 in the polyproline II helix. The next segment corresponds to $n = 1$ and n up to $\frac{1}{3}$ of the chain segment count were considered. The cosines for all trajectory frames of each simulation and each valid n were calculated from which then, the persistence length was derived via Eq. 13.

Dye parameters

see dye-params-table.txt file

Figure S1

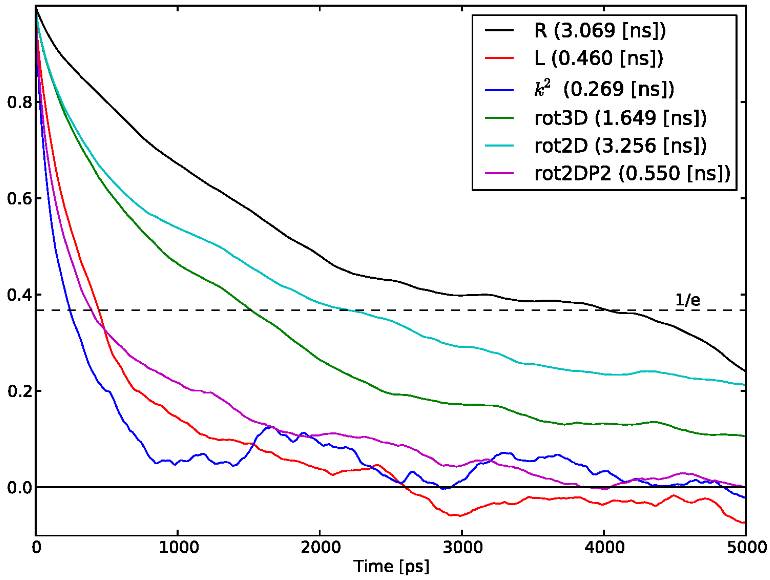


Figure S2

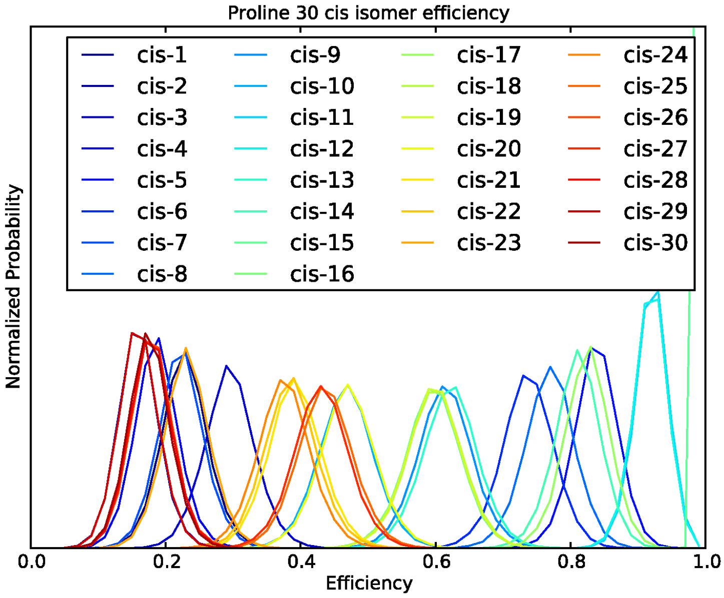


Table S1. Point charge parameterization

Amino acid	OPLS Scaling Factor	StdErr
ALA	0.99	0.06
ARG	1.10	0.08
ASN	0.86	0.13
ASP	1.13	0.15
CYS	0.91	0.14
GLN	0.77	0.14
GLU	0.97	0.09
GLY	0.96	0.15
HIS	1.01	0.13
ILE	0.87	0.09
LEU	0.89	0.06
LYS	0.84	0.13
MET	0.89	0.08
PHE	0.74	0.14
PRO	0.55	0.18
SER	1.02	0.06
THR	0.93	0.10
TRP	0.74	0.15
TYR	0.78	0.14
VAL	0.76	0.14

Comparison between OPLS-AA [57] and DFT point charges of all 20 natural amino acids. Scaling factors with standard errors applied to the DFT charges in order to obtain the best fit for each amino acid. The mean scaling factor is 0.9.

6 Direct Combination of Single-Molecule Förster Transfer and Photoinduced Electron Transfer in Polypeptides from a Comprehensive Analysis of Photon Statistics

In the last chapter of this work I present strategies to combine fluorescent probes for short-range (PET) and long-range (FRET) protein chain dynamics on a single molecule. Novel photon statistical fitting methods to deconvolve the different dynamic contributions, as found in the corresponding ns-FCS measurements, were developed. To verify this novel approach, different test systems of increasing dynamic complexity were produced and extensively studied.

As a starting point, the extent of quenching of the fluorescent dyes Alexa 488 and Alexa 594 by naturally occurring amino acids was explored. Performing ensemble lifetime measurements and ns-FCS measurements allowed us to derive information about the dynamic and static quenching contributions. These results revealed a significant quenching of both fluorescent dyes by Trp which paved the way for further experiments using Alexa 488 and Alexa 594 as a FRET pair and Trp as the quenching amino acid.

Next, polyprolines carrying a single fluorescent dye and a Trp residue at different distances from the labeling position, separated by a flexible Glycine-Serine linker, were used to quantify the extent of the quenching effect at short distances within peptides and proteins. In the next step, a second fluorescent dye was attached to the polyprolines to generate a FRET system consisting of an Alexa 488 dye as a donor and an Alexa 594 as an acceptor fluorescent dye. Highly site specific labeling, and therefore control if either the donor or the acceptor dye was quenched, allowed us to study the effect of donor or acceptor quenching in a static FRET system where the rigidity of the polyprolines between the two fluorescent dyes lead to no observable peptide chain reconfiguration dynamics within the experimental timescales. Extensive studies of the quenching effects which lead to changes in the observed ns-FCS curves, transfer efficiency histograms and ensemble fluorescence lifetimes increased our understanding of the mechanism of these effects and their magnitude.

To complete this work we produced several variants of the N-terminal domain of HIV-1 integrase, an intrinsically disordered protein, to study the convoluted effects of quenching dynamics between a single intrinsic Trp residue and either the donor or acceptor fluorescent dye together with the FRET dynamics between the two fluorescent dyes caused by the diffusional motion of the unfolded polypeptide chain. With this system we were able to show that our novel global fitting strategy for the ns-FCS data, together with constraints from additional fluorescence experiments, allowed us to completely quantify the involved quenching and FRET dynamics. While this study exploits the quenching dynamics to study long and short-range dynamics, the methods presented can also be applied to systems undergoing undesired, and thus complicating, photo physics. The rigorous quantification and separation of these effects, as the one presented here, allows to correct the obtained experimental data and to derive more reliable results from single-molecule experiments.

6.1 Effect of Static and Dynamic Quenching on Gamma and on the Transfer Efficiency

To account for differences in donor and acceptor quantum yields and collection efficiencies, a correction factor γ is commonly used to calculate the transfer efficiency

$$E = \frac{n_A}{n_A + \gamma n_D}, \tag{7}$$

where n_A and n_D are the number of detected fluorescent photons in the acceptor and donor detection channels, respectively. γ is defined as

$$\gamma = \frac{\xi_A Q_A}{\xi_D Q_D}. \tag{8}$$

$\xi_{D,A}$ and $Q_{D,A}$ denote the detection efficiencies and the quantum yields of the fluorescent dyes, respectively. It is calculated from the detection efficiencies of the detectors for the donor and acceptor photons (ξ_D, ξ_A) and from the quantum yields of the used fluorescent dyes (Q_D, Q_A). In the following it will be discussed how E, γ and R_0^6 change due to dynamic and static quenching. We determine the γ of our instruments together with the used fluorescent dyes using pure dye solutions. The obtained γ_o does not include changes in the brightness due to the attachement of the dyes to peptides or proteins.

6.1.1 Dynamic Quenching

Dynamic quenching changes the quantum yield of a fluorescent dye. The new quantum yield, denoted as Q_q, is related to the quantum yield Q_0 in the absence of quenching by

$$Q_q = \frac{\tau_q}{\tau_0} Q_0. \tag{9}$$

τ_q and τ_0 are the mean fluorescent lifetimes in presence and in absence of the quencher. With this equation, the correction factor γ_q, which includes corrections for dynamic donor and acceptor quenching, can be expressed as

$$\gamma_q = \frac{\xi_A Q_{Aq}}{\xi_D Q_{Dq}} = \frac{\xi_A Q_A \frac{\tau_{Aq}}{\tau_A}}{\xi_D Q_D \frac{\tau_{Dq}}{\tau_D}} = \gamma_0 \frac{\tau_D}{\tau_{Dq}} \frac{\tau_{Aq}}{\tau_A}, \tag{10}$$

where γ_0 is the correction factor in the absence of quenching and $\tau_{D,A}$ are the mean donor or acceptor fluorescent lifetimes.

Dynamic donor quenching also changes the Förster radius of the dye pair according to

$$R_0^6 = \frac{9000(ln10)\kappa^2 Q_D J_0}{128\pi^5 n^4 N_A} = \bar{R}_0^6 Q_D. \tag{11}$$

Here the reduced Förster radius $\bar{R}_0^6$ is introduced [91], which is also used in the following equation for the transfer efficiency in presence of a quencher:

$$E_q = \frac{1}{1 + \left(\frac{R^6}{\bar{R}_0^6 Q_D \frac{\tau_{Dq}}{\tau_D}} \right)} \tag{12}$$

Dynamic donor quenching leads to a decrease in transfer efficiency calculated according to equation 7 due to a decrease of the Förster radius.

6.1.2 Static Quenching

In static quenching the quencher can form a nonfluorescent complex with the fluorescent dye. Upon excitation, this complex relaxes radiationless to its ground state (See section 1.1.3). Static donor quenching introduces a quenched donor population, which does not emit photons, nor has the capability to transfer energy from the donor to the acceptor dye due to the very fast relaxation processes within the quenched complex. However, the quantum yield and therefore the Förster radius of the unquenched donor population is not affected by static quenching.

Compared to the timescale of a single molecule diffusing trough the confocal volume, which is typically in the range of 1 ms, the on- and off-rates for the static quenching processes are fast (in the range of hundreds of nanoseconds to microseconds). This leads to an interconversion between quenched and unquenched states of the fluorescent dye during a diffusional event and therefore γ_0 must be adjusted for the reduced amount of acceptor photons. Using the equilibrium constant K_{As} of the static quenching process ($A \leftrightarrows AQ$) it can be written

$$\gamma_q = \frac{\xi_A Q_A \left(1 - \frac{K_{As}}{1+K_{As}}\right)}{\xi_D Q_D} = \gamma_0 \left(\frac{1}{1+K_{As}}\right). \tag{13}$$

The complex formation between the fluorescent dye and the quencher, as found in static quenching, can lead to a change in the absorption spectra and in the extinction coefficient of the fluorescent dye. A change in the absorption spectrum of the acceptor dye changes the overlap integral J_q and hence the Förster radius $R_{0_q}^6$:

$$J_q = \int_0^\infty F_D(\lambda)\varepsilon_{A_q}(\lambda)\lambda^4 d\lambda \tag{14}$$

$$R_{0_q}^6 = \frac{9000(ln10)\kappa^2 Q_D J_q}{128\pi^5 n^4 N_A} \tag{15}$$

The fast interchange between the quenched and unquenched states during the observation time leads to an average transfer efficiency which is given by the equation

$$E_q = \left(\frac{1}{1+K_{As}}\right)\frac{R_0^6}{R_0^6 + R^6} + \left(\frac{K_{As}}{1+K_{As}}\right)\frac{R_{0_q}^6}{R_{0_q}^6 + R^6}. \tag{16}$$

For Alexa 594, the acceptor dye investigated in this study, a slight decrease in the extinction coefficient and a redshift of the absorption spectra was found. This leads to a smaller overlap integral J_q and therefore to a decreased Förster radius compared to the non-quenched system.

6.2 Characterization of the Tryptophan Quenching Effect on Alexa 488 and Alexa 594

Of the naturally occurring amino acids, L-Tryptophan (Trp) is the one with the largest quenching effect on the Alexa 488 and Alexa 594 fluorescent dyes [10]. Therefore we characterized the dynamic and the static quenching interactions of these dyes with free Trp. To obtain information about dynamic quenching, we performed ensemble lifetime experiments at various concentration of Trp, as shown in figure 5(c). Fitting the obtained lifetimes allows us to determine the rate constants for collisional quenching as summarized in table 1. To evaluate whether each collisional encounter leads to dynamic quenching, i.e. whether the dynamic quenching process is diffusion-limited, a calculation of the collision frequency between dye and quencher was performed. The diffusion coefficients of the dyes (including the hydrolyzed linker) and of Trp were calculated using the Wilke-Chang estimation method combined with the additive method by Schroeder for estimating molar volumes at the normal boiling point [87]. The obtained molar volume for Trp is 217 $\text{cm}^3\text{mol}^{-1}$and the corresponding diffusion coefficient in water at 25 °C is $6.7\cdot10^{-6}\text{cm}^2\text{s}^{-1}$, which is in good agreement with experimentally reported values [88, 60]. The calculated values for Alexa 488 and Alexa 594 are listed in table 1. Using an estimated collision radius (R_{col}) of 0.5 nm between the dye and Trp [20, 78] together with the diffusion constants of the dye D_f and the quencher D_q allows us to estimate the collision rate constant (k_{col}) by applying the following equation [60, 20]:

$$k_{col} = 4\pi R\left(D_f + D_q\right) \tag{17}$$

A comparison of the experimental rate constants to those calculated for a diffusion-limited quenching process suggests that the dynamic quenching of Alexa 488 is indeed diffusion-limited. For Alexa 594, the experimentally determined rate constant is significantly smaller than the one expected for a diffusion-limited process.

The static quenching contribution was studied using high-resolution sub-ns FCS experiments with a 1 nM concentration of fluorescent dye and millimolar concentrations of Trp. A typical result of such an experiment at 40 mM Trp with Alexa 488 and Alexa 594 is shown in figure 5(a,b). A significant amplitude caused by the static quenching process was found for both dyes. Fitting these ns-FCS curves using

$$G(\tau) = 1 + a\left(\left(1 - c_{ab}e^{-\frac{\tau}{\tau_{ab}}}\right)\left(1 + c_b e^{-\frac{\tau}{\tau_b}}\right)\left(1 + c_T e^{-\frac{\tau}{\tau_T}}\right)\right) \tag{18}$$

where τ_{ab}, τ_b, and τ_T are the correlation times of photon antibunching, conformational dynamics, and triplet dynamics, c_{ab}, c_b, c_T are the corresponding amplitudes and α is the diffusional amplitude respectively. This fits reveal a larger amplitude (c_b) for Alexa 594 compared to Alexa 488 (See table 1). From the direct relation between this amplitude (c_b) and the equilibrium constant for the static quenching process (K_{Qs}) [20],

$$c_b = \frac{k_{on}}{k_{off}} = K_{Qs} \tag{19}$$

we conclude that Alexa 594 forms more stable complexes with Trp than Alexa 488. As we aim to use static quenching to investigate conformational dynamics within polypeptides, quantitative knowledge of the ratio between the collision rate and the rate for complex formation is needed. Due to the diffusion-limited dynamic quenching constant, a comparison of the collisional quenching rate and the on-rate of the static quenching process can reveal this ratio (See table 1(I)). It is important to mention that this ratio can be influenced by geometrical factors such as the ones introduced by attaching the dyes to a protein chain.

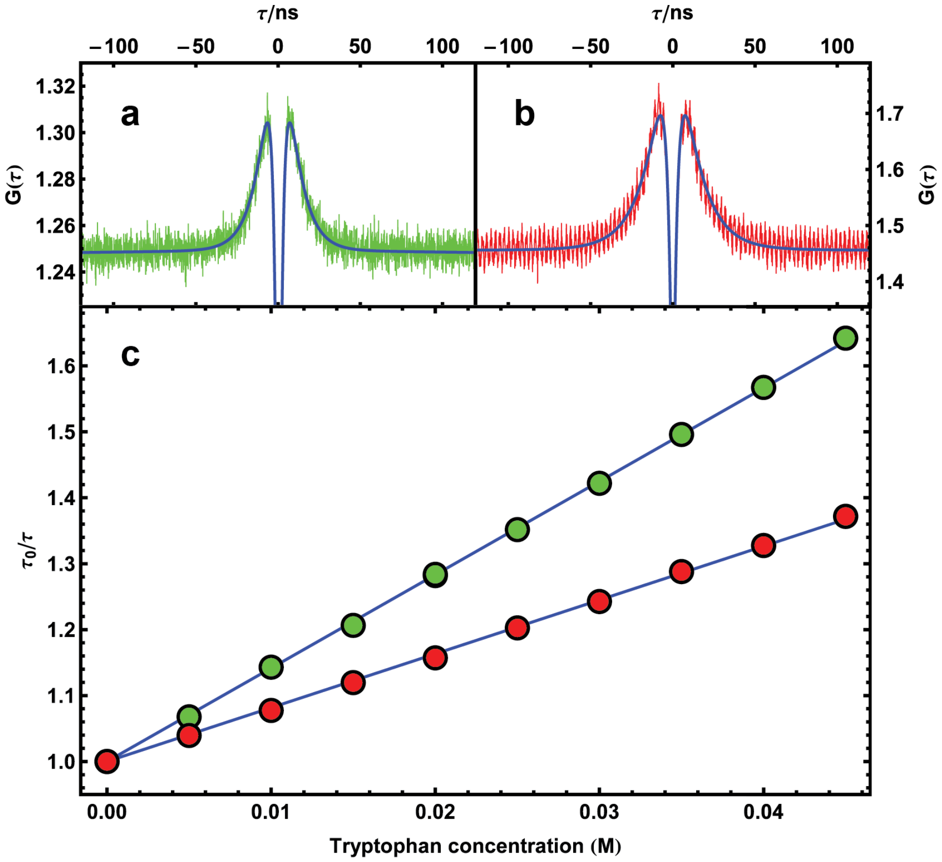


Figure 5: Dynamic and static quenching of Alexa 488 and Alexa 594 with Trp as a quencher. **a,b)** ns-FCS experiments of 1 nM Alexa 488 (green, a)) and Alexa 594 (red, b)) fluorescent dye with 40 mM Trp. The decays on the 10 ns timescale result from the static quenching process. The blue lines indicate the fit using equation 18. The use of a HeNe gas laser, for the excitation of Alexa 594, led to the characteristic ns oscillations superposed on the correlation signal (shown in b). An exchange of the HeNe laser with an identical one had no effect on these oscillations, indicating a general problem introduced by using such lasers. The period of this effect is in the range of 3 ns, indicating a time-of-flight effect within the gas laser tube. **c)** Stern-Volmer plots for Alexa 488 (green) and Alexa 594 (red). From a linear fit of the points using $\frac{\tau_0}{\tau} = 1 + k_q \tau_0 \, [Trp]$, the collisional quenching rate constants shown in table 1 are obtained.

		Alexa 488	**Alexa 594**	
		Ensemble Lifetime Measurements		**%E**
A	Unquenched lifetime	$\tau_0 = 4.06\ \text{ns}$	$\tau_0 = 4.06\ \text{ns}$	2%
B	Stern-Volmer quenching constant	$K_D = 14.2\ \text{M}^{-1}$	$K_D = 8.2\ \text{M}^{-1}$	2%
C	Dynamic quenching rate constant	$k_q = 3.5 \cdot 10^9\ \text{s}^{-1}\text{M}^{-1}$	$k_q = 2 \cdot 10^9\ \text{s}^{-1}\text{M}^{-1}$	5%
D	Quenching rate at 40 mM Trp	$k = 0.14\ \text{ns}^{-1}$	$k = 0.08\ \text{ns}^{-1}$	5%
		ns-FCS Measurements at 40 mM Trp		
E	ns-FCS amplitude	$c_b = 0.68$	$c_b = 1.23$	5%
F	ns-FCS decay time	$\tau_b = 8.53\ \text{ns}$	$\tau_b = 12.50\ \text{ns}$	10%
G	Static quenching on-rate	$k_{on} = 0.048\ \text{ns}^{-1}$	$k_{on} = 0.043\ \text{ns}^{-1}$	20%
H	Static quenching off-rate	$k_{off} = 0.071\ \text{ns}^{-1}$	$k_{off} = 0.036\ \text{ns}^{-1}$	15%
I	Complex formation efficiency	$k/k_{on} = 2.9$	$k/k_{on} = 1.9$	25%
		Calculated Diffusional Parameters		
J	Molecular volume (Schroeder additive)	$V_b = 644\ \text{cm}^3\text{mol}^{-1}$	$V_b = 896\ \text{cm}^3\text{mol}^{-1}$	15%
K	Diffusion coefficient (Wilke-Chang)	$D = 3.5 \cdot 10^{-6}\ \text{cm}^2\text{s}^{-1}$	$D = 2.9 \cdot 10^{-6}\ \text{cm}^2\text{s}^{-1}$	25%
L	Diffusion coefficient 2-Focus FCS	$D = 3.7 \cdot 10^{-6}\ \text{cm}^2\text{s}^{-1}$	-	10%
M	Collision rate constant (Equation 17)	$k_{col} = 3.9 \cdot 10^9\ \text{s}^{-1}\text{M}^{-1}$	$k_{col} = 3.6 \cdot 10^9\ \text{s}^{-1}\text{M}^{-1}$	65%

Table 1: Results from quenching experiments of Alexa 488 and Alexa 594 with Trp as a quencher. In the right column relative errors are shown. In the upper part, the quenching rates extracted from ensemble lifetime experiments of fluorescent dyes for different concentrations of Trp are shown. **A)** Lifetimes of the unquenched fluorescent dyes in buffer (100 mM KPO4, 143 mM Mercaptoethanol, 0.001% Tween 20, pH 7.2). **B)** Dynamic quenching constants from Stern-Volmer plots shown in figure 5 c). **C)** Dynamic quenching rate constants calculated using $k_q = K_D/\tau_0$. **D)** For comparison with the ns-FCS experiments (See figure 5 (a,b)), the dynamic quenching rate at 40 mM Trp are shown. In the second part of the table, the results for the ns-FCS measurements performed at 40 mM Trp are shown. **E/F)** Amplitudes and decay times from the fits to the ns-FCS data shown in figure 5 (a,b). **G/H)** On-rates and off-rates of the static quenching complexes calculated using a two-state quenching model and therefore the following relationship: $c_b = \frac{k_{on}}{k_{off}}$ and $\tau_b = \frac{1}{k_{on}+k_{off}}$ [20]. **I)** Assuming a diffusion-limited dynamic quenching process, the ratio between the dynamic quenching rates at 40 mM Trp (D) and the on-rate of the static quenching processes (G) reports on the number of dye-Trp encounters needed until a statically quenched complex is formed successfully. In the last part of the table, the results for the calculations of the diffusion coefficients and of the theoretical collision rate constants are shown. **J)** Molecular volume at the normal boiling point of the dyes including the hydrolyzed linker calculated using Schroeder‘s additive method [80, 87]. **K)** Diffusion coefficient calculated by the Wilke-Chang estimation method using the molecular volume shown in J) [80, 87]. **L)** For comparison the diffusion coefficient of Alexa 488 was measured using a 2-focus FCS setup (Bengt Wunderlich, personal communication). **M)** The theoretical collision rate constant was calculated using equation 17.

As discussed in section 6.1.2, the interaction between a static quencher and the acceptor fluorescent dye can lead to a change in the absorption spectra and therefore in R_0. In order to quantify this effect, absorption spectra of Alexa 594 were measured at various concentrations of Trp (See figure 6). From these spectra, the overlap integrals and the change in R_0 was calculated (See figure 7)[1]. At the solubility limit of Trp in aqueuos buffer (approximately 50 mM), convergence was not yet achieved for the values of R_{0q}. Therefore the points were fitted with a binding isotherm (blue line in figure 7):

$$R_0\left(Trp\right) = \frac{K_{sq} R_0\left(0\right) + R_{0q}\left[Trp\right]}{K_{sq} + \left[Trp\right]} \tag{20}$$

From this fit, it was possible to assign values to the equilibrium constant (K_{sq} =25.5 mM), the unquenched Förster radius (R_0 =54.1 Å) and the quenched Förster radius (R_{0q} =52.0 Å). A comparison to the equilibrium constant obtained from the ns-FCS experiments which is 33.5 mM, showed a good agreement.

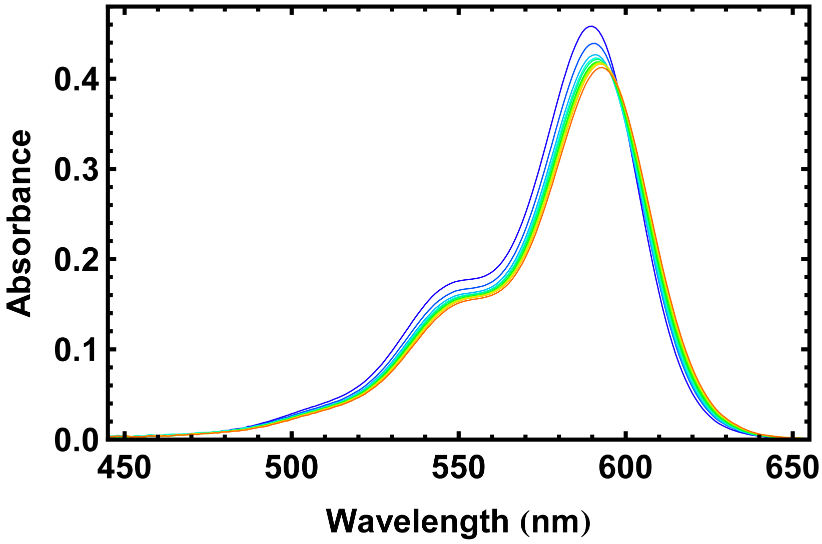


Figure 6: Absorption spectra of Alexa 594 recorded at different concentrations of Trp (0-50 mM, Blue → Red). For Alexa 594 static quenching leads to a redshift of the absorbance spectra together with a reduction of the extinction coefficient. A similar behavior was reported for the oxazine derivative MR121 [20].

---

[1] The largest error in calculating R_0 results from the uncertainty in the extinction coefficient of the Alexa 594 fluorescent dye. Since this dye is usually bought from Invitrogen in milligram quantities, there is not enough material available to determine the extinction coefficient. That is the reason why we heave to rely on the values provided by the manufacturer. Unfortunately the specified values fluctuated between 73'000 and 92'000 M^{-1} cm^{-1} during the last years.

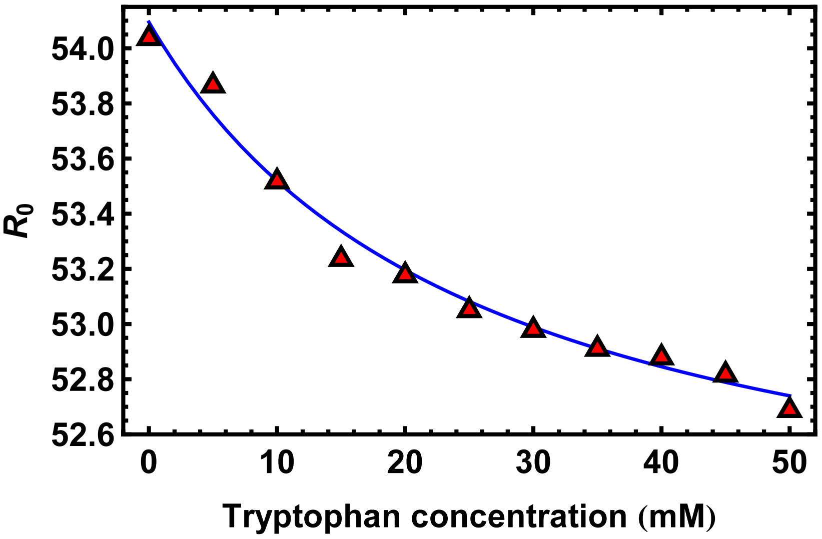


Figure 7: Förster radius at different concentrations of Trp. The blue line is a fit to the data according to equation 20. From the experimental scatter and the uncertainty in the extinction coefficient of Alexa 594, the estimated error is 5%.

With these experiments, all relevant parameters for characterizing the quenching of Alexa 488 and of Alexa 594 with Trp were obtained. Both dyes are dynamically and statically quenched. For Alexa 488, we showed that the dynamic quenching process is diffusion-limited. These experiments demonstrate the possibility of using the Alexa 488/594 FRET pair with Trp as a quencher in a single experimental system. Theoretically, the planarity of the fluorescent dyes would allow simultaneous stacking of two Trp residues on separate sides of the delocalized electron system. However, due to no Trp concentration dependence in the experimentally found off-rates of the static quenching reaction, such a scenario is improbable for the Alexa 488 and the Alexa 594 fluorescent dyes.

6.3 Polyprolines as a System to Study Quenching Effects on a Static FRET Distance

6.3.1 Design and Synthesis of Polyproline Variants

The polyproline constructs were designed building on the experience from earlier experiments on a similar system performed by Schuler *et al.* [96]. For the Alexa 488 and the Alexa 594 fluorescent dyes as FRET probes for single-molecule experiments, polyprolines with 20 proline residues as rigid spacers between the fluorescent dyes were shown to lead to a mean transfer efficiency of approximately 0.6 [96]. In order to have a high sensitivity for FRET changes together with an approximately equal amount of donor and acceptor photons, this spacer distance was chosen for the new constructs. In order to introduce PET quenching, a flexible (Glycine-Serine)$_n$ linker, carrying the quenching Trp, was attached on the C-terminal end of the polyproline. Neuweiler *et al.* used the oxazine fluorophore MR121, attached to similar linkers, to study distance-dependent quenching effects of Trp [78]. They found a strong quenching effect for linker lengths of 4-10 residues. Therefore we decided to make constructs with similar distances spanning a range from 2 to 8 residues between the cysteine to which the fluorescent dye was attached and the Trp. In order to avoid effects on the transfer efficiency histograms due to varying steric constraints introduced by different linker lengths, the same length was used for all peptides with Trp located at different positions in the linker. Figure 8 gives an overview of the different peptide constructs. The corresponding sequences together with the labeling positions are summarized in table 9. In addition to the constructs and the data presented in this work, new variants of the fluorescent dyes with different linker lengths were synthesized and attached to the same peptide constructs. This allowed us to obtain insights into the influence of the dye-linker length on quenching dynamics.

The peptides were produced by solid phase synthesis and purified to 80-90% according to analytical HPLC and ESI-MS[2]. The reduced cysteines were either labeled with Alexa Fluor 488C_5-maleimide[3] or Alexa Fluor 594C_5-maleimide[4] according to the protocol supplied by the manufacturer. An initial purification was performed using size exclusion chromatography[5]. A further purification step using reverse phase chromatography[6] was performed to remove residual dye molecules. A typical HPLC gradient from 0.1 % TFA in H_20 to 40% Acetonitrile over 50 minutes was performed. The flow rate was at 1 ml/min. The gradients were optimized for each peptide by using multiple analytical runs. The lyophilized peptides were redissolved in a 200 mM pH 9.2 sodium bicarbonate buffer and labeled on the N-terminal glycine using either Alexa Fluor 488 succinimidyl ester[7] or Alexa Fluor 594 succinimidyl ester[8]. The labeling of the glycine residue was found to be difficult. Various attempts using different solvents, incubation times, temperatures and sonication steps of the samples during the labeling reactions were done. Sonicating the samples during the reaction together with an up to 50 fold molecular excess of reactive dye proved to be the method of choice for the second labeling step. The large excess of dye prohibited a direct purification of the constructs via size exclusion or reverse phase columns due to potential irreversible contamination of the columns. Therefore the amount of free dye was first depleted using disposable columns such as gravity flow[9] or spin[10] columns. Multiple cycles of reverse phase purification followed this initial step. The labeling success was controlled using single-molecule spectroscopy and ESI-MS.

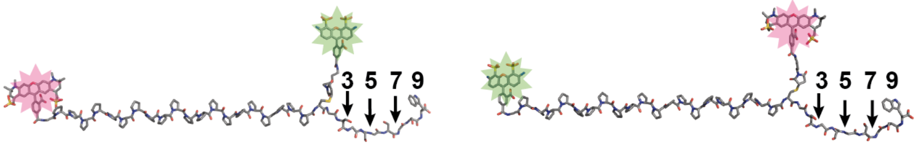


Figure 8: Overview of the polyproline constructs investigated in this work. On the left side a construct with a quenched donor and on the right side with a quenched acceptor is shown. The fluorescent dyes are highlighted by green (Alexa 488) and red (Alexa 594) stars. The numbers indicate the different positions of the quenching Trp. In the structures shown here, the Trp is at the position called W9.

---

[2] PEPTIDE 2.0, Chantilly, VA, USA

[3] A10254, Invitrogen, Eugene, Oregon, USA

[4] A10256, Invitrogen, Eugene, Oregon, USA

[5] Superdex Peptide 10/300GL, GE Healthcare, UK

[6] XTerra RP18 Column, 5 µm, 4.6 x 250 mm, 186000496, Milford, MA, USA

[7] A20000, Invitrogen, Eugene, Oregon, USA

[8] A20004, Invitrogen, Eugene, Oregon, USA

[9] Polyacrylamide Desalting Columns, 1800, 5 ml, 1862253, Thermo Scientific, Rockford, IL, USA

[10] Zeba Spin Desalting Columns, 7K MWCO, 0.5cml, 89882, Thermo Scientific, Rockford, IL, USA

Peptide variant	Sequence
	1 11 21 31
PP20REF	**G**PPPPPPPPP PPPPPPPPPP P**C**GSGSGSGS A594 A488 A488 A594
PP20W3	**G**PPPPPPPPP PPPPPPPPPP P**C**GS**W**GSGSG S A594 A488 A488 A594
PP20W5	**G**PPPPPPPPP PPPPPPPPPP P**C**GSGS**W**GSG S A594 A488 A488 A594
PP20W7	**G**PPPPPPPPP PPPPPPPPPP P**C**GSGSGS**W**G S A594 A488 A488 A594
PP20W9	**G**PPPPPPPPP PPPPPPPPPP P**C**GSGSGSGS **W** A594 A488 A488 A594

Figure 9: Sequence overview of the polyproline constructs with the labeling positions. The two labeling positions consisting of the N-terminal glycine and of a cysteine residue at position 22, as well as the quenching Trp, are indicated in bold. Below the labeling positions, the different variants, where either the donor (Alexa 488) or the acceptor (Alexa 594) dye is quenched, are indicated. For simplicity, the peptide names are abbreviated (Ref, W3, W5, W7, W9) in the following discussions.

6.3.2 Experimental Results

Fluorescence Lifetimes Ensemble fluorescence lifetimes were measured using a custom-built fluorescence lifetime spectrometer, mainly consisting of a supercontinuum fiber laser[11] and a microchannel photomultiplier tube[12] in a perpendicular configuration, connected to a photon counting module[13]. Depending on the available amount of sample, the measurement concentration was either 0.1 or 1 µM in a standard measurement buffer (See section 6.3.2). To avoid contributions by rotational effects, all the lifetimes were measured in a magic angle configuration where the fluorescence emission was measured at an angle of 54.7 relative to the excitation polarization. Before and after each measurement series, an instrument response function (IRF) was recorded. The resulting data was fitted using a fitting function consisting of the IRF and of a single exponential decay term. The data from all constructs was well fitted by a single exponential decay, indicating a homogeneous sample composition. Figure 10 shows two fluorescence lifetime measurements as an example. A significant difference can be seen between the polyproline with (dark green) and the one without Trp (green).

---

[11] SC-450-PP, Fianium, Southampton, UK

[12] PMT R3809U-50, Hamamatsu, Shizuoka, Japan

[13] PicoHarp 300, PicoQuant GmbH, Berlin, Germany

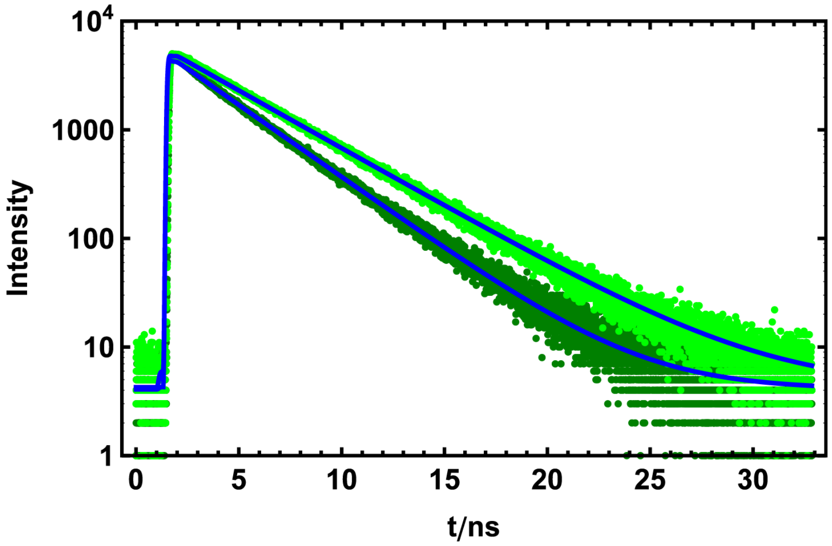


Figure 10: Example fluorescence lifetime measurements of donor only polyproline peptides. The bright green points show an experiment with the Ref construct, lacking a Trp. The dark green points show a measurement with the W3 construct. The blue lines indicate the corresponding fits.

Figure 11 shows the results for the lifetime fits. Comparing the values for the reference ("Ref") to the values for the constructs containing a quenching Trp (3-9) reveals a significant dynamic quenching contribution, which is, within the accessible distance space, almost independent of the sequence separation between the dye and the quencher. As for the experiments with free dye and free Trp (See section 6.2), a larger dynamic quenching effect was found for Alexa 488 compared to Alexa 594. From assuming that the dynamic quenching observed for the labeled peptides is a reporter for collisional encounters between the dye and the quencher as for the free dye (See section 6.2), and therefore for the chain dynamics of the flexible linker, its possible to conclude that there is only a small observable distance effect for the separations used here. This finding is in good agreement with the results reported by Neuweiler et al. where they found almost identical quenching on-rates for dye-quencher separations of 4-10 amino acids [78]. As a possible explanation they mention the finite dimension of the dye-quencher reporter system [78]. Using triplet-triplet energy transfer as a reporter, Krieger et al. also measured contact rates of glycine-serine polypeptides [58]. They found a Gaussian chain behavior for peptides with distances longer than 20-30 amino acids. Below this value, they found a reduced distance scaling and convergence of the contact formation rates. This effect was explained as an increase in chain stiffness at short distances due to specific intrachain interactions and excluded volume effects. To verify that only the dye on the cysteine (see table 9) is dynamically quenched, the lifetimes of variants carrying an Alexa 488 dye on the cysteine and an Alexa 594 on the non-quenched N-terminus were measured using direct excitation of the Alexa 594 dye. As shown in figure 11, no significant difference in the lifetimes between the reference and the constructs containing a Trp was found. Therefore its safe to state that there is indeed no dynamic quenching effect on the N-terminal labeling position.

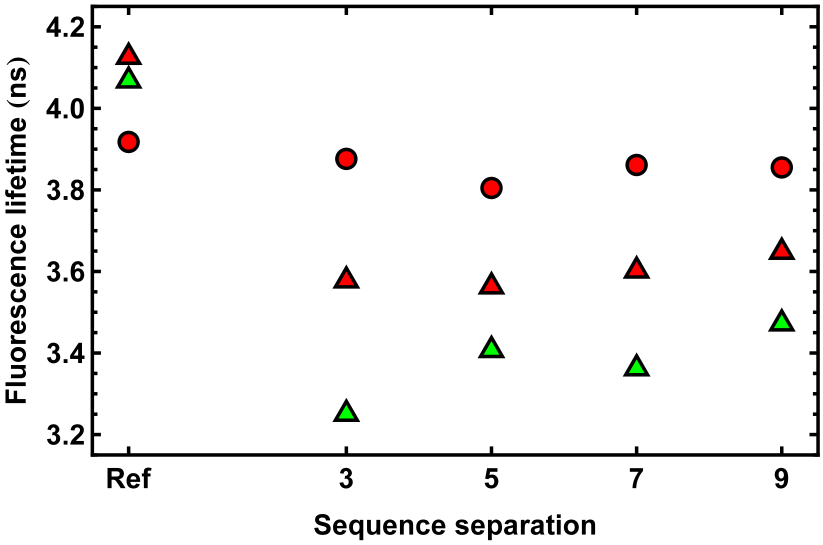


Figure 11: Results from single exponential fits of ensemble lifetimes measured in a magic angle configuration to avoid contributions by rotational effects. Red triangles: Acceptor quenched variant measured using direct acceptor excitation. Green triangles: Donor-only variants of the donor quenched constructs. Red circles: As a control, acceptor direct excitation of the donor quenched variants. Different sequence separations between the fluorescent dye and the quenching Trp are indicated on the x-axis. "Ref" represents the lifetimes measured for the corresponding control lacking a Trp. The errors are within the size of the symbols.

Histograms and Transfer Efficiencies Single-molecule experiments were carried out for all constructs using 50 pM labelled sample in 100 mM Potassium Phosphate buffer, 143 mM Mercaptoethanol, 0.001% (v/v) Tween 20, pH 7.2. The laser power of the 488 nm CW laser was adjusted to 100 µW. The usual measurement time was one hour. Fluorescent bursts were detected using a lower threshold of 30 photons per event. To reduce the effect of photobleaching, bursts with an assymetric distribution of red and green photons were removed from the analysis. As discussed in section 6.1, a correction factor γ is used to calculate the transfer efficiency histograms from the raw photon counts of a single-molecule experiment. However, the use of a single correction factor γ does not allow to rigorously correct for all the influencing factors. Therefore we implement this γ correction using a matrix approach which introduces corrections for differences in the fluorescence quantum yields of the free dyes, detection efficiencies for green and for red photons, direct excitation of the acceptor as well as crosstalk between the collection channels. The relation between the raw photon counts $\mathrm{n}_{\mathrm{A},0}$ and $\mathrm{n}_{\mathrm{D},0}$, for example those measured in a two channel confocal microscope, and the corrected values $\mathrm{n'}_{\mathrm{A}}$ and n_{D} can be expressed by the following matrix equation [39]:

$$\begin{pmatrix} n_{A,0} \\ n_{D,0} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} n'_A \\ n_D \end{pmatrix} + \begin{pmatrix} b_A \\ b_D \end{pmatrix} \tag{21}$$

The matrix a_{ij} consists of the cumulative correction factors for differences in quantum yields, the different collection efficiencies for donor and acceptor photons, and the crosstalk between the detectors. The matrix consisting of b_{A} and b_{D} contains the background count rates for each detection channel. Subtracting the background counts, followed by a multiplication by the inverse of the a_{ij} matrix allows calculation of the corrected raw counts:

$$\begin{pmatrix} n'_A \\ n_D \end{pmatrix} = \left(\begin{pmatrix} n_{A,0} \\ n_{D,0} \end{pmatrix} - \begin{pmatrix} b_A \\ b_D \end{pmatrix} \right) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} \tag{22}$$

The experimental correction factors $\mathrm{a_{ij}}$, which are unique parameters of the confocal instrument together with the used fluorescent dyes, are measured using micromolar solutions of fluorescent dyes (With hydrolized reactive groups). Attaching the dyes to proteins or peptides, as in the case of the polyprolines shown here, can lead to dynamic and static quenching and therefore to a change in the correction factors. Knowing the change of the fluorescence lifetimes for the donor and the acceptor dyes upon quenching (τ_{QD}, τ_{QA}) together with the knowledge of the static quenching constant of the acceptor dye (K_{As}) allows the calculation of the corrected raw photon counts by using the following equation:

$$\begin{pmatrix} n'_A \\ n_D \end{pmatrix} = \left(\begin{pmatrix} n_{A,0} \\ n_{D,0} \end{pmatrix} - \begin{pmatrix} b_A \\ b_D \end{pmatrix} \right) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} (1+K_{As})\frac{\tau_{0A}}{\tau_{QA}} & 1 \\ 1 & \frac{\tau_{0D}}{\tau_{QD}} \end{pmatrix} \tag{23}$$

Here the $\mathrm{a_{ij}}$ values are obtained by using pure dye solutions with the corresponding fluorescence lifetimes (τ_{0A}, τ_{0D}). As a last step, n'_A has to be corrected for direct excitation of the acceptor according to

$$n_A = n'_A - (n'_A + n_D)\varepsilon_A/(\epsilon_A + \epsilon_D), \tag{24}$$

where ϵ_A and ϵ_D are the extinction coefficients of donor and acceptor at the excitation wavelength.

Figure 12 shows the resulting transfer efficiency histograms obtained for the polyproline peptides with quenched donor. The column on the left shows the analysis performed using the $\mathrm{a_{ij}}$ values from the dye calibration. Introducing a Trp (W3-W9) leads to a shift of the calculated transfer efficiencies due to the change in the donor lifetime and therefore in the corresponding donor quantum yield. Measuring the ensemble fluorescence lifetimes of donor quenched donor-only constructs and using the values to perform a correction according to equation 23, results in the calculated transfer efficiency histograms shown on the right column. Here, the only difference between the reference construct and the constructs carrying a Trp should arise from the quenching induced change in the Förster radius. The multiplication by the quenching correction (See equation 23) leads to an artificial increase of the number of photons and therefore to an increase in the number of found bursts as indicated by the numbers on the left side of the histograms.

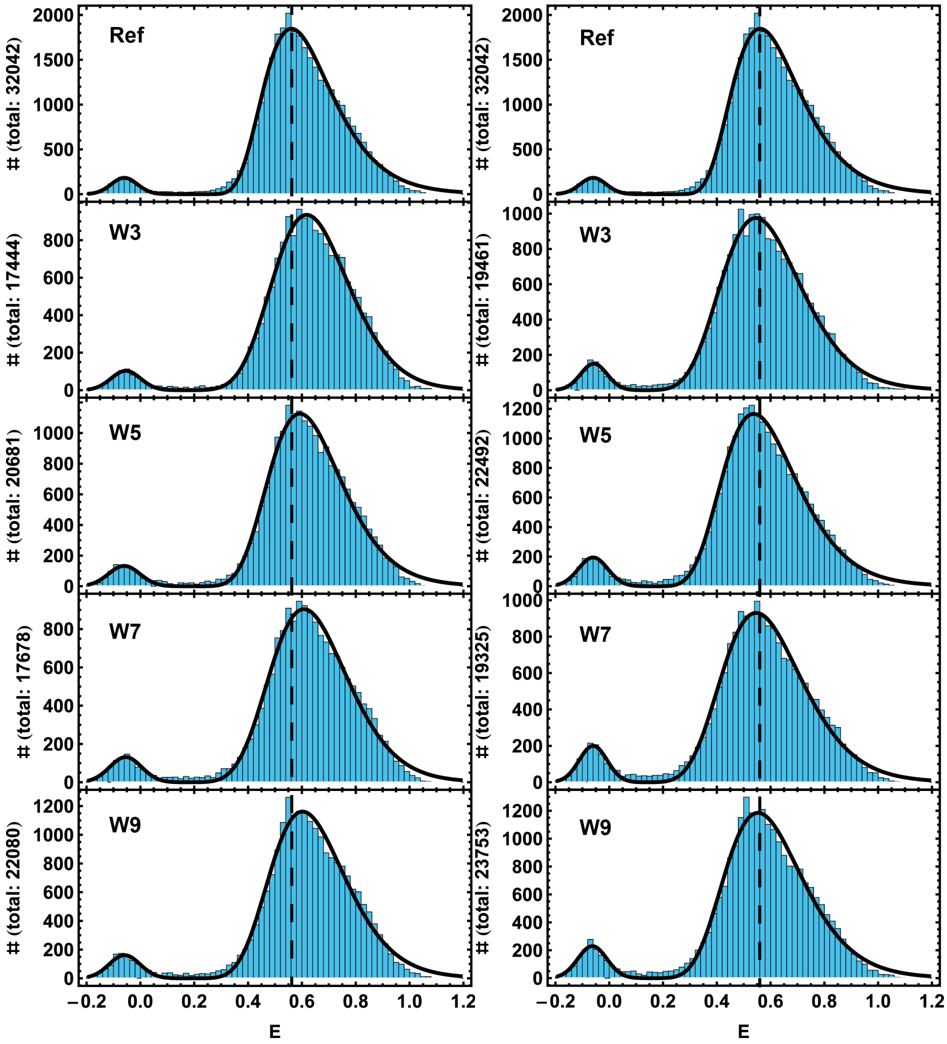


Figure 12: Transfer efficiency histograms of the polyproline peptides with quenched donor. Each row shows histograms with an increasing distance separation between the donor dye and the quenching Trp as indicated by the inset. As a reference, a construct lacking Trp is included (Ref). The black line tracing the histogram indicates the fit of the two populations, in which the donor-only peak at zero transfer efficiency was fitted with a Gaussian and the transfer peak with a Lognormal distribution. The left column shows the histograms without correction for the quenching effects. The right column shows histograms from the same experiments including the matrix countrate corrections (See equation 23) for dynamic quenching effects. The dashed black line is a guide to the eye to highlight the offset relative to the reference measurement. The y-axis indicates the number of detected fluorescent events (bursts).

Figure 13 shows the polyproline peptides with quenched acceptor. In contrast to donor quenching (figure 12), the multiplicative effect of dynamic and static acceptor quenching leads to a much more dramatic effect if no quenching correction is applied. The left column again shows the calculated histograms using the correction values from a pure dye calibration. Introducing the Trp (W3-W9) leads to a change of up to 0.2 in the calculated transfer efficiencies. Applying the correction using equation 23 with the ensemble fluorescence lifetimes measured by acceptor direct excitation together with the static equilibrium quenching constants derived from the static quenching on- and off-rates of the photon statistics fits (See section 6.3.3), leads to the histograms shown in the right column which are in good agreement with the reference construct. An alternative strategy to measure the static acceptor quenching is by performing ns-FCS experiments using acceptor direct excitation and fitting the data with a two state quenching model (See caption of table 1).

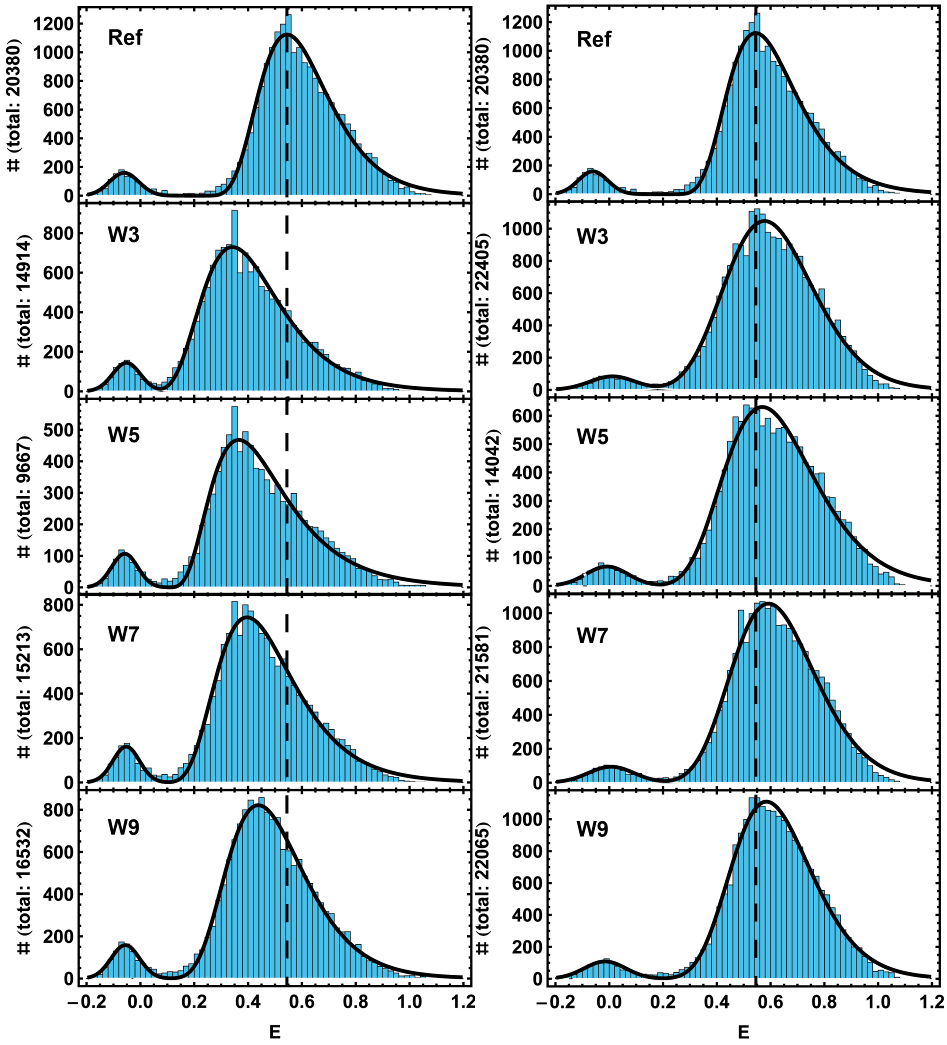


Figure 13: Transfer efficiency histograms of the polyproline peptides with quenched acceptor. Each row corresponds to an increasing distance separation between the acceptor dye and the quenching Trp as indicated by the inset. As a reference, a construct lacking a Trp is included (Ref). The black line tracing the histogram indicates the fit of the two populations, in which the donor-only peak at zero transfer efficiency was fitted with a Gaussian and the transfer peak with a Lognormal distribution. The left column shows the histograms without correction for the quenching effects. The right column shows histograms from the same experiments including the matrix countrate corrections (See equation 23) for dynamic and static (photon statistics fits of ns-FCS data) quenching effects. The dashed black line is a guide to the eye to highlight the offset relative to the reference measurement. The y-axis indicates the number of detected fluorescent events (bursts).

To summarize the influence of Trp quenching on the calculated transfer efficiency histograms and to underscore the importance of applying the correct correction factors, the peaks in the transfer efficiency histograms were fitted using a Lognormal distribution and the resulting transfer efficiencies are shown in figure 14. This illustrates the strong effect of quenching on the transfer efficiency histograms. In the most prominent case, namely in the acceptor quenching of the W3 peptide, the quenching leads to a shift of around 0.2 in the observed transfer efficiency. It also shows, how an exact quantification of the quenching effects together with the application of the introduced correction strategy can again lead to reliable results.

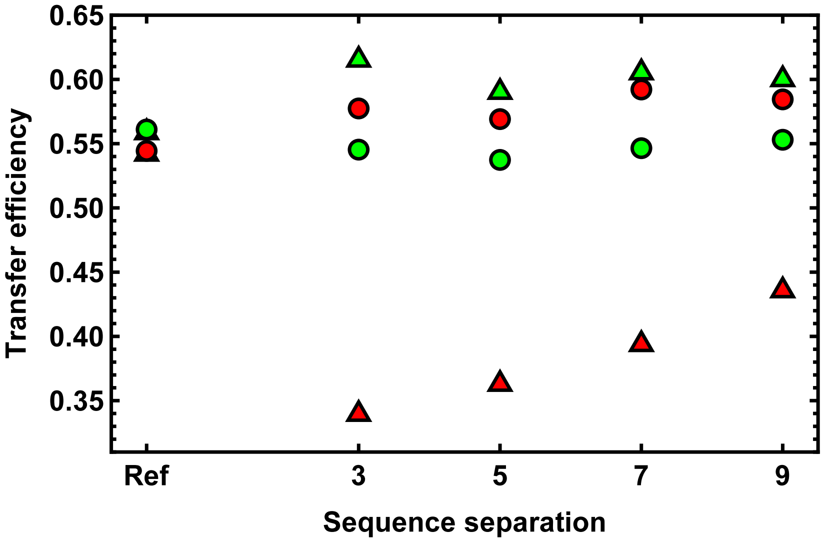


Figure 14: Transfer efficiencies of the polyproline constructs obtained by fitting the transfer peak of the calculated transfer efficiency histograms using a Lognormal distribution. Plotted is the maximum of the Lognormal distribution. Different sequence separations between the fluorescent dye and the quenching Trp are indicated on the x-axis. “Ref” represents the transfer efficiency measured for the control constructs lacking Trp. The green symbols show the values for donor quenching and the red symbols for acceptor quenching. The triangles show the transfer efficiencies calculated using a default matrix countrate correction and therefore no correction for quenching effects. The circles show the transfer efficiencies calculated using a matrix countrate correction which is corrected for static and dynamic quenching effects. The error in fitting transfer efficiency histograms is approximately 5%.

6.3.3 Using Photon Statistics to Probe the Quenching Dynamics and to Explain the Experimental ns-FCS Data

Starting with the first ns-FCS measurements performed in our group, we have used an elegant theoretical formalism introduced by Irina Gopich and Attila Szabo for calculating theoretical intensity correlation curves [27, 29, 74, 75]. The formalism allows incorporating both photophysical effects such as photon antibunching and triplet blinking as well as the effect of the conformational dynamics of the polypeptide chain via FRET in the photon statistics calculation of ns-FCS curves [27, 29].

However, in the earlier studies, the calculated ns-FCS curves were not directly fitted to the experimental data. Instead, the experimental curves were fitted with a much simpler model as shown in equation 18 [74, 75]. From all fit parameters, only τ_b was compared to the corresponding value obtained from the theoretical calculations. A quantitative comparison of the amplitudes was not included. This approach is justified by the assumptions that, first, the timescales of all three dynamic processes captured in the above equation are well separated (antibunching: 1-5 ns, chain dynamics: 20-100 ns, and triplet blinking: 1-10 µs), and second, that within experimental error, all processes are sufficiently well described by single exponential correlation functions. This simplifying assumptions hampered a meaningful quantitative comparison of the amplitudes, and hence a full understanding of the experimental curves. However, it could be verified that the obtained time constant τ_b was correct and this was the main focus of the earlier work [74].

In the cases studied in this work, where we aim to disentangle FRET dynamics and PET dynamics, which occur simultaneously and on comparable timescales, it becomes crucial to understand the measured ns-FCS curves in much more detail. Hence, we decided to fit the calculated correlation curves directly to the experimental ns-FCS data and we extended the underlying model by the previously neglected process of singlet-singlet annihilation. Of course also the effect of donor and acceptor quenching by Trp (or other fluorescence quenchers) had to be included as well as more technical effect such as detection crosstalk, acceptor direct excitation, etc.

In the following sections I will first give a general description of the mathematical method used for calculating fluorescence correlation functions for a complex photophysical system. Then I will introduce the specific models used for the systems studied here. I will further give some technical information on how to calculate the ns-FCS curves efficiently and finally I will present the result of the analysis of the experimental data.

Calculating Intensity Correlation Functions from Kinetic Models In order to calculate the correlation functions, the formalism developed by Gopich *et al.* was used [29]. A rate matrix $\mathbf{K}$ describes all electronic and conformational transitions between the states of the system [74]. The off-diagonal elements of the matrices $\mathbf{V}_\mathrm{D}$ and $\mathbf{V}_\mathrm{A}$ contain the radiative rates of the monitored transitions [74]. The time dependence of the system is described by a population vector $\mathbf{p}(t)$ which solves the rate equation:

$$\frac{d\mathbf{p}(\mathbf{t})}{dt} = \mathbf{K}\mathbf{p}(t), \tag{25}$$

where the sum of all elements of $\mathbf{p}$ is normalized to one ($\mathbf{1}^\top \mathbf{p} = 1$). With $\mathbf{p}_\mathrm{SS}$ we denote the steady state solution of equation 25, i.e. $\mathbf{K}\mathbf{p}_{SS} = 0$. The fluorescence correlation functions can now be calculated by [74, 28]:

$$G_{ij}(\tau) = \frac{\mathbf{1}^\top \mathbf{V}_\mathrm{j} e^{\mathbf{K}\tau} \mathbf{V}_\mathrm{i} \mathbf{p}_{SS}}{(\mathbf{1}^\top \mathbf{V}_\mathrm{i} \mathbf{p}_{SS})\,(\mathbf{1}^\top \mathbf{V}_\mathrm{j} \mathbf{p}_{SS})} \text{ with } i, j = A, D \tag{26}$$

where $\mathbf{1}$ is the unity vector, τ is the delay time and $e^{\mathbf{K}\tau}$ is the matrix exponential of $\mathbf{K}\tau$. Correlation functions with $i = j$ are called autocorrelations and for $i \neq j$ crosscorrelations.

Photon Statistics Model Implementation A major challenge in applying such a photon statistics approach is to find a suitable model adequately describing the experimental system without introducing too many states that are inaccessible to experimental investigation. A four-state FRET model consisting of a non fluorescent ground state (DA), a donor- (D*A) and an acceptor-excited state (DA*) as well as a double-excited state (D*A*) was used as a starting point (see middle section of figure 15). This model was extended to introduce static donor and acceptor quenching. For static donor quenching, two new states were introduced, DQA and DQA*. In both states, the donor quencher complex DQ is formed, whereas the acceptor is in its ground state A or excited state A*. A* can be excited by the laser light with the rate constant k_exA. From the absorption spectra of Alexa Fluor 488 and 594 one finds $k_{exA} \simeq 0.05 \cdot k_{exD}$ at the excitation wavelength of 488 nm. Analogously, one introduces two states DAQ and D*AQ for describing acceptor quenching. It is important to take into account that static quenching leads to a change in the absorption spectra of the acceptor and therefore to a changed FRET transfer rate due to the change in Förster radius (See section 6.2). $k_{FQ}(r)$ is the corrsponding changed Förster transition rate constant, which describes the depopulation of D*AQ to DAQ. This effect leads also to a significant change in the donor-donor autocorrelation function $G_{DD}(\tau)$.

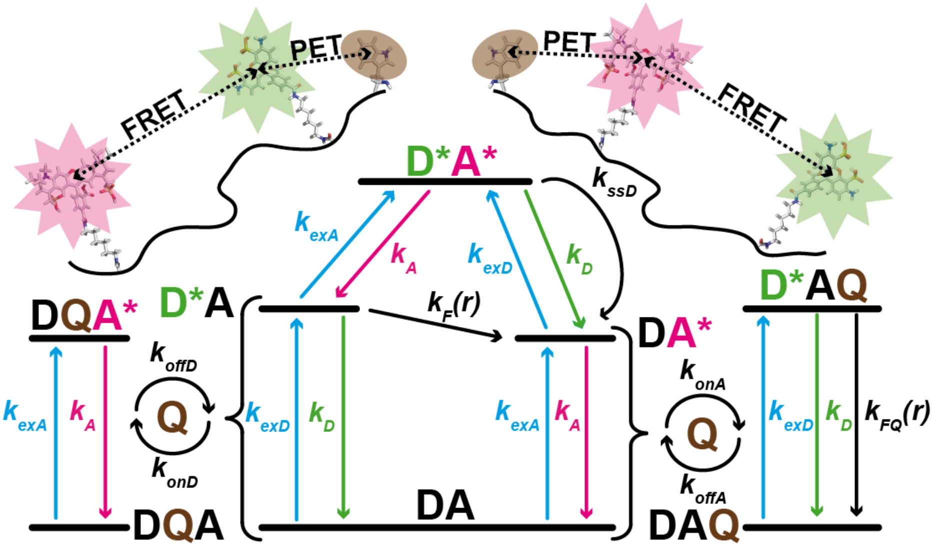


Figure 15: Jablonsky diagram of the electronic and molecular transitions of a FRET system in presence of static donor and acceptor quenching. In the upper corners illustrations of the dye and quencher arrangement are shown. Both FRET systems consist of an Alexa 488 dye as a donor (green star) and an Alexa 594 dye (red star) as an acceptor. The PET quencher Trp is highlighted in brown. On the upper left side, the situation for donor quenching is shown and on the upper right side for acceptor quenching. The systems were designed such that only one dye, either the donor or the acceptor, is quenched by the TRP.

The experimental systems are designed such that either the donor or the acceptor dye is quenched by the Trp residue. The second fluorescent dye is sufficiently far away from the quencher that its quenching can be neglected.

In the case of donor quenching, the following six states are needed: DA, DQA, DQA*, D*A, DA*, D*A* (same order as in the matrices). The rate matrix describing the electronic and conformational transitions is

$$\mathbf{K} = \begin{pmatrix} K_{11} & k_{offD} & 0 & k_D & k_A & 0 \\ k_{onD} & K_{22} & k_A & k_{onD} & 0 & 0 \\ 0 & k_{exA} & K_{33} & 0 & k_{onD} & k_{onD} \\ k_{exD} & 0 & 0 & K_{44} & 0 & k_A \\ k_{exA} & 0 & k_{offD} & k_F & K_{55} & k_D + k_{SSD} \\ 0 & 0 & 0 & k_{exA} & k_{exD} & K_{66} \end{pmatrix}, \tag{27}$$

where the diagonal elements K_{ii}, are given by the requirement that the sum over the elements of each column is zero. The radiative transition matrices are given by $\mathbf{V}_D = \xi_D Q_D k_D \tilde{\mathbf{V}}_D$ and $\mathbf{V}_A = \xi_A Q_A k_A \tilde{\mathbf{V}}_A$, where ξ_i and Q_i are the detection efficiency and the quantum yield, respectively. $\tilde{V}_D$ and $\tilde{V}_A$ are given by

$$\tilde{\mathbf{V}}_D = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tilde{\mathbf{V}}_A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{28}$$

In the case of acceptor quenching, the system consists of the following states: DA, DAQ, D*A, DA*, D*AQ, D*A*. The rate matrix describing the electronic and conformational transitions is

$$\mathbf{K} = \begin{pmatrix} K_{11} & k_{offA} & k_D & k_A & 0 & 0 \\ k_{onA} & K_{22} & 0 & k_{onA} & k_D + k_{FQ} & 0 \\ k_{exD} & 0 & K_{33} & 0 & k_{offA} & k_A \\ k_{exA} & 0 & k_F & K_{44} & 0 & k_{SSD} + k_D \\ 0 & k_{exD} & k_{onA} & 0 & K_{55} & 0 \\ 0 & 0 & k_{exA} & k_{exD} & 0 & K_{66} \end{pmatrix}. \tag{29}$$

$\tilde{V}_D$ and $\tilde{V}_A$ are given by

$$\tilde{\mathbf{V}}_D = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tilde{\mathbf{V}}_A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{30}$$

Note that since the prefactors $\xi_i Q_i k_i$ of the radiative transition matrices cancel in equation 26, we can write:

$$G_{ij}(\tau) = \frac{\mathbf{1}^\top \tilde{\mathbf{V}}_{\mathrm{j}} e^{\mathbf{K}\tau} \tilde{\mathbf{V}}_{\mathrm{i}} \mathbf{p}_{SS}}{\left(\mathbf{1}^\top \tilde{\mathbf{V}}_{\mathrm{i}} \mathbf{p}_{SS}\right)\left(\mathbf{1}^\top \tilde{\mathbf{V}}_{\mathrm{j}} \mathbf{p}_{SS}\right)} \tag{31}$$

Effect of Detection Channel Crosstalk on the Correlation Functions Due to the broad emission spectrum of Alexa Fluor 488, some donor photons are detected in the acceptor detection channel. We define the crosstalk β as the ratio between donor photons detected in the acceptor detection channel and donor photons detected in the donor channel. For our instrument, the crosstalk was determined to be $\beta \simeq 0.08$. In order to calculate more accurate correlation functions, we have to take the crosstalk into account. This is done by modifying the detection rate matrix for the acceptor detection channel according to

$$V_{A_C} = V_A + \beta V_D. \tag{32}$$

Dividing this equation by $\xi_A Q_A k_A$ we obtain

$$\tilde{V}_{A_C} = \tilde{V}_A + \beta \cdot \frac{1}{\gamma} \cdot \frac{k_D}{k_A} \cdot \tilde{V}_D. \tag{33}$$

Effect of the donor-only population on the correlation functions Due to imperfections in the purification as well as due to photo bleaching of the fluorescent dyes, our samples contain subpopulations of polypeptides lacking an active donor or acceptor dye. These subpopulations contribute to the fluorescence signal and in consequence also to the ns-FCS curves. Here we consider only the subpopulation of peptides which have an active donor but not an active acceptor fluorescent dye. We call it the 'donor-only population'. The subpopulation of peptides having only an active acceptor dye is neglected as the 488 nm laser light only weakly excites it.

The influence of the donor-only population on the correlation function can be described by using the general FCS formula for a mixture of two species. Denoting the donor-only population by d and the population where both donor and acceptor are active as f, we obtain:

$$G_{i,j}(\tau) = \frac{N_f B_{f,i} B_{f,j} G^f_{i,j}(\tau) + N_d B_{d,i} B_{d,j} G^d_{i,j}(\tau)}{\left(N_f B_{f,i} + N_d B_{d,i}\right)\left(N_f B_{f,j} + N_d B_{d,j}\right)} \tag{34}$$

Here N_f and N_d are the mean number of particles of each species present in the confocal volume, $B_{f,i}$ and $B_{d,i}$ are the relative brightnesses of the species as observed in detection channel i, and $G^f_{i,j}$ and $G^d_{i,j}$ are the correlation functions for each species, where again i and j denote the detection channels. For the relative brightnesses one finds:

$$B_{f,D} = (1-E)\,, B_{f,A} = \gamma(E+\alpha) + \beta\,(1-E)\,, B_{d,D} = 1, \text{ and } B_{d,A} = \beta \tag{35}$$

Here the brightness of the donor-only species detected in the donor detection channels is set to one. Further we introduce the ratio $r_d = N_d/N_f$ and we finally obtain for the correlation functions:

$$G_{DD}\,(\tau) = \frac{(E-1)^2\,G^f_{DD}\,(\tau) + r_d G^d_{DD}(\tau)}{N_f(1-E+r_d)^2} \tag{36}$$

$$G_{AA}\,(\tau) = \frac{r_d\beta^2 G^d_{AA}\,(\tau) + (\beta - E\beta + (E+\alpha)\,\gamma)^2\,G^f_{AA}(\tau)}{N_f\,((1-E+r_d)\,\beta + (E+\alpha)\,\gamma)^2} \tag{37}$$

$$G_{DA}\,(\tau) = \frac{r_d\beta G^d_{DA}\,(\tau) + (E-1)((E-1)\beta - (E+\alpha)\gamma)G^f_{DA}(\tau)}{N_f(-1+E-r_d)((-1+E-r_d)\beta - (E+\alpha)\gamma)} \tag{38}$$

$$G_{AD}\,(\tau) = \frac{r_d\beta G^d_{AD}\,(\tau) + (E-1)((E-1)\beta - (E+\alpha)\gamma)G^f_{AD}(\tau)}{N_f(-1+E-r_d)((-1+E-r_d)\beta - (E+\alpha)\gamma)} \tag{39}$$

This approach can be extended for additional subpopulations within an equilibrium ensemble such as a high FRET native state population in equilibrium with an unfolded population of the same molecule.

Effect of Triplet Blinking on the correlation functions Another effect which must be included into the calculation of the correlation functions, is triplet blinking. In contrast to ideal fluorescent systems, the fluorescent probes used in these experiments do have the tendency to populate long-lasting triplet states, directly affecting the correlation functions. In the process of this work, complex kinetic schemes were developed, which also included triplet states for the donor and for the acceptor dye. For including these states, new rate constants must be introduced, such as quenching rates for the triplet states and FRET rates to/from the triplet states of the donor and the acceptor molecule. Besides the increasing computational complexity, these rates are difficult to measure and there is the risk of overparametrisation. Nevertheless, the laser powers used in these experiments lead to observable triplet blinking in the microseconds regime. As this is at least one order of magnitude slower than the rates found for PET and FRET dynamics, we have separation of timescales, and we hence can write for the correlation function which includes PET and FRET dynamics as well as triplet blinking:

$$G_{ij}(\tau) = 1 + c_{ij}\left(1 + c_{T_{ij}} e^{-\frac{\tau}{\tau_{T_{ij}}}}\right) G_{IJ}(\tau). \tag{40}$$

The correlation functions calculated using photon statistics and the combining functions shown above (G_{IJ}, see equations 36 - 39) can hence be expanded by an additional exponential term consisting of a triplet amplitude $\left(c_{T_{ij}}\right)$ and a triplet decay $\left(\tau_{T_{ij}}\right)$ time as shown in equation 40. An additional global amplitude (c_{ij}) is needed to include effects such as varying concentrations of fluorescent molecules and diffusion.

Fitting the Experimental Data The nummerical calculations of the correlation functions were done using Wolfram Mathematica[14]. To perform global fits to measured donor-donor, donor-acceptor and acceptor-acceptor correlations, required to dissect the contributing factors such as PET and FRET dynamics, time efficient Mathematica code was needed. The following list highlights some aspects in programming the fitting algorithm:

- **Numerical Calculations:** As default, Mathematica tries to solve problems symbolically. Strict enforcement of numerical variable handling reduced the needed time per calculation.
- **Caching:** Only a part of the fitting process needs to invoke time consuming processes such as the eigendecomposition of the rate matrix $\mathbf{K}$. The local optimization of the triplet decay time or the global amplitude as an example is a very fast process. Caching the results of subcalculations and checking if the needed parameters have changed since the last execution lead to a substantial increase in performance
- **Compilation of time critical functions:** Using the built in Compile[] function either with the built-in compiler, or more difficult, using an external C-compiler lead to significant speed improvements. Due to the time needed for the compilation, this only is efficient for functions which are called very often.
- **Parallel computing:** Parallelization was implemented either on the level of a single measurement where each of the correlations (DD,DA,AA) was calculated on a separate Mathematica subkernel or on several measurements where up to 24 datasets were fitted in parallel.
- **No redundancy:** The code was checked to avoid calculating the same value more than once.
- **Benchmarking:** For every command in the code the performance was checked using the AbsoluteTiming[] command and the best performing implementation was chosen.

These optimizations made the global fitting of the ns-FCS data practical. Numerous fitting strategies with different sets of free parameters could be evaluated. An increasing robustness of the fitting results was found by fixing as many parameters as possible. Single-molecule together with ensemble fluorescence experiments allow the precise determination of a variety of parameters needed in the photon statistics fits. Table 2 shows the provenance of the fixed parameters and the corresponding range found for different constructs or measurements. As shown in equation 11, dynamic donor quenching does change the quantum yield of the donor and therefore the Förster radius. This rises the questions whether donor quenching leads to a change in the energy transfer rate constant (k_F). The rate of energy transfer can be written using the following equation:

$$k_F = k_D \frac{R_0^6}{r^6} \tag{41}$$

Introducing dynamic quenching leads to a change in k_D and in R_0^6:

$$k_{FQ} = k_D \frac{\tau_D}{\tau_{DQ}} \frac{R_0^6 \frac{\tau_{DQ}}{\tau_D}}{r^6} \tag{42}$$

From this follows that k_F and k_{FQ} are identical since the change in k_D and the change in donor quantum yield cancel.

---

[14] Wolfram Mathematica 8, Wolfram Research, Champaign, IL, USA

Fixed Parameters	**Value and Method of Determination**
Excitation rate constant ($k_{exD}, k_{exA} = \alpha k_{exD}$)	Antibunching of Trp-free donor-only construct: 0.02 ns^{-1}, 0.001 ns^{-1}[74]
Ground state relaxation rate constant (k_D, k_A)	Ensemble lifetime measurements: 0.24-0.3 ns^{-1}
Energy transfer rate constant (k_F, k_{FQ})	Transfer efficiency of Trp-free construct together with corrections for R_{0Q}: 0.23 – 0.29 ns^{-1}
Relative donor-only population (r_d)	Relative histogram peak areas (donor-only / transfer population) of corresponding E histograms: 3.5-7%
Förster radii (R_0, R_{0Q})	Dye emission and absorption spectra and extinction coefficients: 54.1Å, 52.1Å
Triplet state decay times (τ_{ij})	Exponential fits of Trp-free control constructs: 2.7 - 3.4 µs
Crosstalk (β)	Calibration measurements: 8 %
Fitting Parameters: Time origins ($t_{0_{ij}}$), Overal amplitudes (c_{ij}), Quenching rates (k_{onQ}, k_{offQ}), Annihilation rate constant (k_{SSD}), Triplet amplitudes ($c_{T_{ij}}$)	-

Table 2: Parameters for fitting the correlation functions and how they were determined. The first column (Fixed Parameters) declares the values needed by or derived using the photon statistics fit. The second column (Value and Method of Determination) shows the origin of the fixed parameters. In cases where precise values were measured for each construct, the corresponding range is given. In the last row, the fitting parameters are shown.

The correlation data of all available constructs were fitted using identical values for initializing the fit procedure. Therefore no bias was introduced by manually adjusting the fit parameters. To weight the autocorrelations and the crosscorrelation equally, the chi-square value of the crosscorrelation was multiplied by a factor of two.

Figure 16 shows the ns-FCS data together with the fits of the peptides with donor quenching, figure 17 the ns-FCS for the acceptor quenching. Each row in these figures shows the correlations for one construct in the order $G_{DD}(\tau)$ (green), $G_{DA}(\tau)$ (orange), and $G_{AA}(\tau)$ (red). The result of the global fit is shown in blue. As expected for the stiff polyproline system, no dynamic component can be seen in this case. Therefore the only amplitudes in addition to fluorescence antibunching and the triplet state are introduced by the static quenching of the dye by Trp.

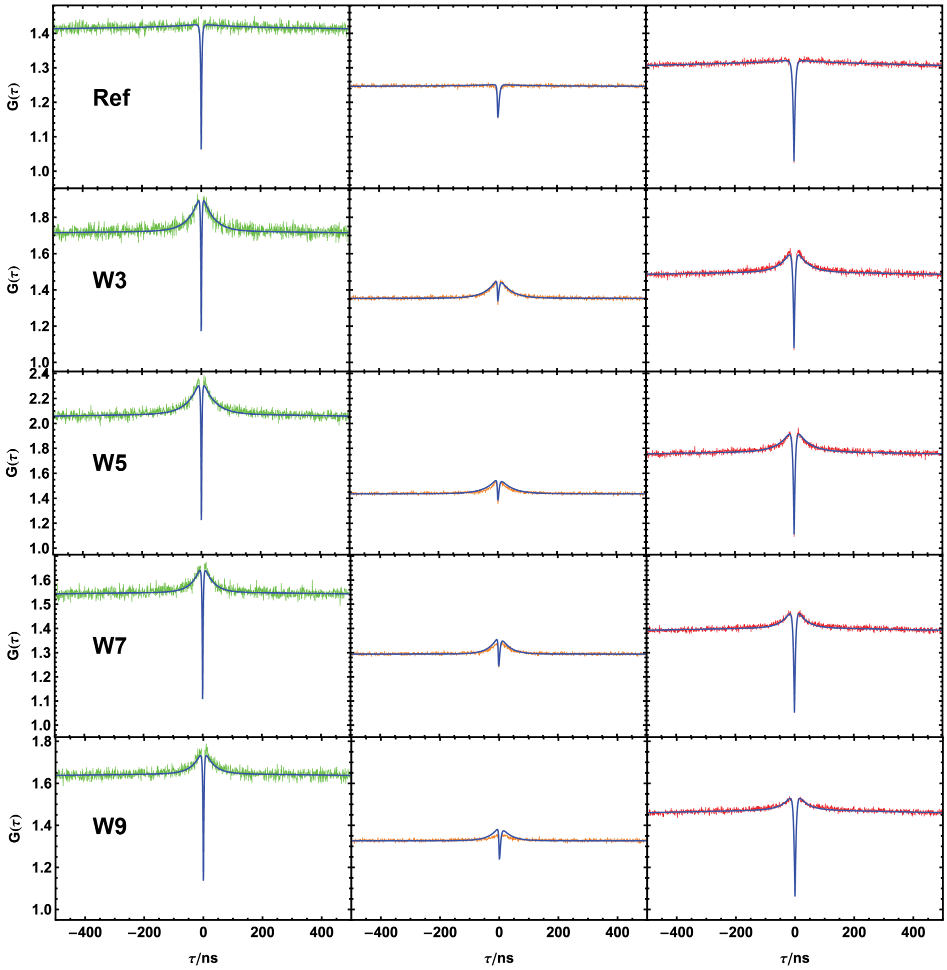


Figure 16: Overview of the ns-FCS measurements together with the global photon statistics fits of polyproline constructs with quenched donor. Each row shows the donor-donor (green), donor-acceptor (orange) and the acceptor-acceptor (red) ns-FCS of a single experiment. The blue line shows the fit result from the photon statistics fits, globally fitting all three correlations. The label in the left column indicates the sequence separation between the donor dye and the quenching Trp residue. The first row, labeled with "Ref", presents the measurement of an identically labeled control construct lacking a Trp. For a structural overview of the constructs see figure 8.

Whereas donor quenching leads to very prominent quenching amplitude in all three correlations, acceptor quenching mostly affects the acceptor-acceptor correlation. Additionally, the change in R_0 upon quenching of the acceptor leads to a small but observable effect in the donor-donor and the donor-acceptor correlation as seen in figure 17, W3. In general, with increasing distance between the fluorescent dye and the quenching Trp, a decrease in the quenching amplitude can be found.

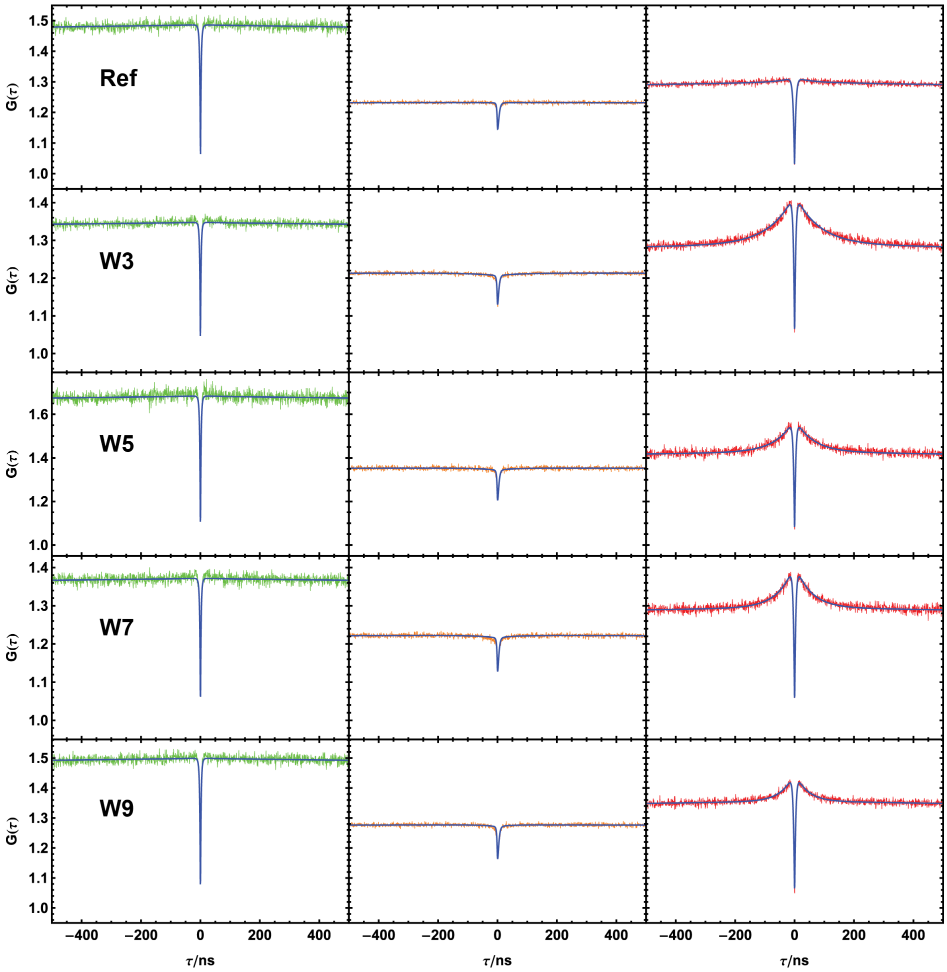


Figure 17: Overview of the ns-FCS measurements together with the global photon statistics fits of polyproline constructs with quenched acceptor. Each row shows the donor-donor (green), donor-acceptor (orange) and the acceptor-acceptor (red) ns-FCS of a single experiment. The blue line shows the fit result from the photon statistics fits, globally fitting all three correlations. The label in the left column indicates the sequence separation between the acceptor dye and the quenching Trp residue. The first row, labeled with “Ref”, presents the measurement of an identically labeled control construct lacking a Trp. For a structural overview of the constructs see figure 8.

To highlight the decrease in the amplitudes with increasing distance separation between the dyes and the quencher, all the acceptor-acceptor correlations of the polyprolines with quenched donor or quenched acceptor are shown in the same plot (See figure 18(a,b)). The quenching rates extracted from the photon statistics fits are shown in figure 18(c). As already shown for the ensemble lifetimes, the on-rate is largely independent of the sequence separation between the fluorescent dye and the quencher. However, for the off-rates, a clear and very similar trend is found for both dyes. With increasing distance separation between the dye and the quencher, an increase in the off-rates was found. This indicates that increasing the linker length between the two decreases the stability of dye-quencher complexes. A possible explanation are steric effects introduced by the additional residues between the dye and the quencher. The relative offset in the off-rates between the two different dyes with a lower stability of the Alexa 488 - Trp complexes is similar to the values found for free dyes and Trp (See table 1H).

The following formula can be used to calculate the quenching rate from the ensemble lifetimes:

$$k_q = \frac{1}{\tau_q} - \frac{1}{\tau_0} \tag{43}$$

Assuming a diffusion-limited quenching process for the Alexa 488 dye, this rate should be identical to the contact rate between the dye and the Trp. For an unquenched lifetime of 4.08 ns and the lifetime of the W3 construct of 3.26 ns, the quenching and therefore the contact rate should be 0.062 ns^{-1}. A comparison to the corresponding value from the photon statistic fit, which is 0.0074 ns^{-1}, reveals a 8.4 fold difference of the two rates. According to the calibrations performed using free dye and Trp, this difference should be around 2.9 (See table 1I). Therefore, attaching Alexa 488 and the quenching Trp to a peptide chain seems to reduce the probability of a successful static complex formation upon collisional contact by a factor of around 3. Introducing a peptide chain seems to affect the collisional quenching encounters differently than the ones leading to a successful static complex formation. Calculating the probability of a successful static complex formation upon collisional contact for the bulkier Alexa 594 dye results in an even higher value of around 4 indicating the possibility of a significant steric contribution in this effect. However, it is important to mention that the theoretical calculation revealed a not fully diffusion-limited dynamic quenching process for Alexa 594 (See table 1) and the errors in these factors may be rather large and therefore caution should be used when assigning the relevance to a change in this factor from 3 to 4.

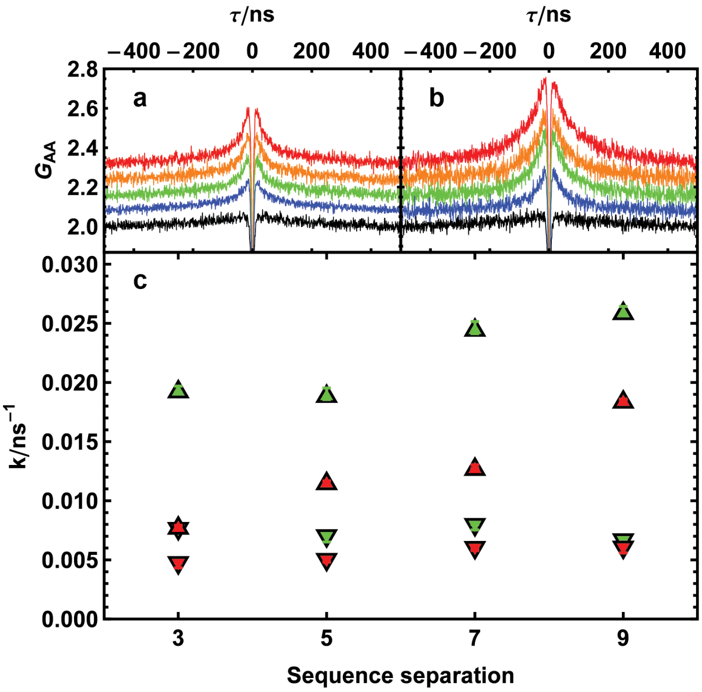


Figure 18: Changes in amplitudes and quenching rates from photon statistics fits. **a,b)** Acceptor-acceptor ns-FCS of donor (a) and of acceptor (b) quenching. To avoid a contribution of the donor-only-population, for both constructs the acceptor-acceptor correlation is shown. Color-coding: red (W3), orange (W5), green (W7), blue (W9), black (Reference construct without Trp). For clarity, the plots are normalized and presented with an equidistant offset of 0.08. **c)** On- and off-rates resulting from the photon statistics fits of the quenched constructs with increasing sequence separation between the dye and the quenching Trp. The green triangles represent the values for donor and the red for acceptor quenching. Triangles pointing down show the on-rate of the static quenching process and the ones pointing up, the off-rates. The error bars resulting from independent measurements of the same constructs are within the size of the symbols.

The experiments on the polyproline constructs revealed the interplay between dynamic and static quenching together with FRET at a fixed distance. It was shown how various distances between the dye and the quencher do affect the dynamic and the static quenching components. By measuring single-molecule transfer efficiency histograms, the change in the calculated transfer efficiencies upon introduction of an additional quencher were shown. Strategies to quantify these effects and to correct the experimental data were introduced. Comprehensive ns-FCS measurements allowed the development of a quantitative photon statistics model to finally fit the experimental data in a global fashion and therefore, for the first time, using all the available experimental information. A comparison of the probabilities to form a static complex upon a quenching encounter revealed a significant difference between free dye and quencher in contrast to systems where the dye and the quencher are attached to a polypeptide. This may indicate a significant steric contribution in these probabilities, which has to be considered in deriving mechanistical conclusions from experiments using static quenching as a reporter system. The successful implementation and description of quenching together with a static FRET system paves the way for the next level of complexity, namely the combination of a dynamic FRET system together with an additional quencher.

6.4 The N-Terminal Domain of the Intrinsically Disordered Domain of HIV-1 Integrase as a Model System for Quenching and FRET Dynamics

6.4.1 Design, Expression and Labeling of the N-Terminal Domain of HIV-1 Integrase

In order to have PET and FRET quenching in a single molecule, we used a modified variant of a previously developed N-terminal domain of HIV-1 integrase (IN) which is unfolded in the absence of Zn^{2+} ions [72]. HIV-1 integrase is a 290 amino acid protein which is responsible for the insertion of viral sequences into the host DNA [37]. It consists of three structural domains, a 50 amino acid N-terminal domain with a zinc binding motif, a 160 amino acid central domain thought to be the catalytic center of the protein and an 80 amino acid long C-terminal domain needed for dimerization or multimerization [37]. Cai *et al.* solved the solutions structure of the dimeric N-terminal domain of HIV-1 Integrase (residues 1-55) using NMR [7]. They found two interconverting conformations differing in the coordination of zinc by two histidines. These two binding modes result in large conformational differences [7]. Nomura *et al.* conducted NMR studies on a single point mutant (Y15A) of the N-terminal domain which showed correct folding upon Zn^{2+} binding but only one conformational state was populated [81]. In order to work with a homogenous system, the same mutation was introduced into our sequence (Y19A in the sequence shown in table 19). A recent study showed that the N-terminal Zn^{2+} binding domain plays a key role in sequence-specific cooperative DNA binding [8].

Protein variant	Sequence						
	1	11	21	31	41	51	60
IN 488W_594	GSHMFLDGID	**C**AQEEHEKAH A488	SN**W**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD	**C** A594
IN 488F_594	GSHMFLDGID	**C**AQEEHEKAH A488	SNFRAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD	**C** A594
IN 594W_488	GSHMFLDGID	**C**AQEEHEKAH A594	SN**W**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD	**C** A488
IN 594F_488	GSHMFLDGID	**C**AQEEHEKAH A594	SNFRAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD	**C** A488

Figure 19: Overview of the different IN variants. The two labeling positions, consisting of C11 and C60, together with the quenching Trp (W23) are indicated in bold. Additional known quenchers of Alexa 488 are underlined [10]. The fluorescent dyes are indicated below the labeling positions. On the left, the construct abbreviations, used in the further discussion, are introduced.

Initially we designed the sequence of integrase to measure short-range quenching dynamics within the cavity of the GroEL/ES chaperone system. A single cysteine residue was introduced at position 11, hence allowing site-specific labeling with a fluorescent dye. The plasmid containing the coding sequence was produced using custom gene synthesis[15]. Introduction of Atto Oxa11[16] as a fluorescent reporter revealed a strong quenching effect of the dye at the C11 position by the intrinsic Trp residue on position W23, confirming the general suitability of the system as a quenched probe [127]. The IN variants used in this study were expressed and purified by Franiska Zosel. Using site-directed mutagenesis, an additional cysteine was introduced at position 60. Additionally, a control construct was produced with the W23 exchanged to F23 and therefore removing the PET quencher. Table 19 shows the various IN sequences together with the labeling positions and the abbreviations used troughout. The integrase genes were cloned into a pET15b vector, expressed and the initial purification was performed on a Ni-NTA agarose column[17] using the cleavable N-terminal hexahistidine tag (His-Tag). After thrombin[18] cleavage of the His-Tag, a second Ni-NTA affinity chromatography was performed to remove uncleaved IN. In the final step, the cleaved material was further purified using high-resolution reverse phase chromatography[19]. Following lyophilization and solvation, the molecular mass was confirmed by mass spectrometry. Labeling with both the donor and the acceptor fluorescent dyes was performed using the introduced cysteines together with reactive malemide groups on the fluorescent dyes. Homogeneous, site-specific labeling is an obligatory requirement for the deconvolution of FRET and PET quenching. A significant effort was invested to achieve site-specific labeling without having an orthogonal chemical labeling strategy. The combination of the different chemical reactivities of the C11 and the C60 for the two different dyes together with extensive purifications using reverse phase chromatography allowed the separation of the two different labeling isomers, and therefore site specific labeling with purities of around 95% was achieved. As a complication, the IN variant used here carries a Zn^{2+}-binding moiety consisting of two histidines and two cysteines, providing additional possibilities for labeling at unwanted reactive positions. To avoid the labeling of these cysteines, the labeling reactions were conducted in presence of 0.5 mM $ZnCl_2$, which induces folding due to complex formation between Zn^{2+} and the intrinsic cysteines and two

---

[15] Celtek Biosciences, Nashville, TN, USA

[16] ATTO-TEC, Siegen, Germany

[17] Ni-NTA Agarose, 30230, Qiagen, Düsseldorf, Germany

[18] Thrombin from bovine plasma, 36402, SERVA Electrophoresis, Heidelberg, Germany

[19] XTerra RP18 Column, 5 µm, 4.6 x 250 mm, 186000496, Milford, MA, USA

histidines, therefore removing the chemical reactivity of these cysteines. The first labeling reaction was performed using a substochiometric ratio between the Alexa 594 dye and IN followed by reverse phase purification and separation of the C11 and the C60 labelled isomers. Next, the purified isomers were labeled using the Alexa 488 fluorescent dye in equimolar concentrations. After the second reverse phase purification, the mass of the doubly labeled IN was confirmed using mass spectrometry. To verify the identity of the labeling positions, a tryptic digest was performed and the mass of the resulting fragments was analyzed using mass spectrometry. A representative structure of the unfolded IN 488W_594 is shown in figure 20.

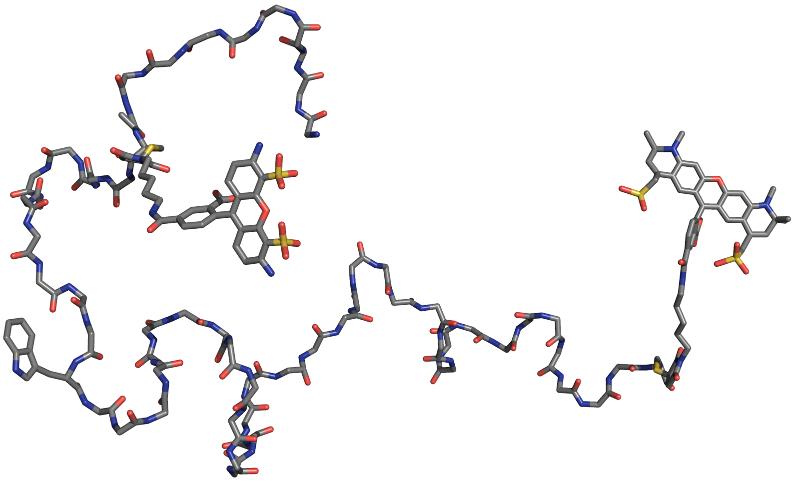


Figure 20: Structural overview of the unfolded IN 488W_594 variant (See table 19). The backbone representation shows the quenching Trp and the quenched Alexa 488 fluorescent dye on the left side. On the right the acceptor dye Alexa 594 is shown.

6.4.2 Experimental Results

In the following section, the experimental results for the investigated IN variants are presented. GdmCl concentration dependent measurements were performed with the IN variants in order to probe additional dynamic processes. Adding GdmCl to the intrinsically disordered IN affects the expansion and the dynamics of the protein chain and therefore the FRET distance and the PET quenching in multiple ways. Due to its high net charge at physiological pH, IN is rather expanded as a result of charge repulsion at low ionic strength [72]. Adding GdmCl, and therefore increasing the ionic strength, leads to a screening of the charged amino acids and therefore to a compaction of the unfolded chain. However, at increasing concentrations, GdmCl leads to a reexpansion of the chain [72]. In addition to the expansion, the observed FRET dynamics are affected by the change in solvent viscosity upon increasing GdmCl concentrations. For the observed PET quenching, the change in the unfolded chain dimension together with increasing solvent viscosity upon addition of GdmCl can lead to changes in the observed on-rates. The denaturing effect of GdmCl also affects the stability of the static dye-quencher complex and therefore leads to a change in the on- and off-rates for the static quenching.

Fluorescence Lifetimes Fluorescence lifetimes were measured as described for the polyprolines (See section 6.3.2). To modulate the unfolded chain dimension of the IN variants and to weaken the quenching interaction between the quenching Trp and the fluorescent dye, GdmCl dependent measurements were performed. Figure 21(a) shows the results for the integrase constructs with a quenched donor. The differences between the lifetimes of IN 488W (green triangles) and IN 488F (green circles) highlight the dynamic quenching effect introduced by Trp. Increasing the GdmCl concentration reduces the difference between the two constructs, indicating a reduction in the dynamic quenching effect. Most probably this is due to the change in solvent viscosity. Direct acceptor excitation of IN 488W_594 (red triangles) in contrast to IN 488F_594 (red circles) shows almost no difference in the fluorescence lifetime. This indicates that there is no significant dynamic quenching effect between the W23 and the acceptor dye on the C-terminal C60 position. Figure 21(b) shows the lifetime results for the integrase with a quenched acceptor dye. In contrast to Figure 21(a), a difference in the acceptor lifetimes (red) between IN 594W_488 (triangles) and IN 594F_488 (circles) variants can now be seen. This is a clear indication of dynamic quenching of the acceptor dye on the C11 position by the Trp on position 23. A rather unexpected result was the difference in the donor lifetimes for the constructs with a quenched acceptor dye. From the large sequence separation between the donor dye on C60 and the Trp on position W23, no quenching effect was expected. The observed difference points to a much stronger dynamic quenching effect for the Alexa 488 dye compared to Alexa 594. This is in good agreement with the results obtained using free dye and quencher (See section 6.2). A photon statistics description of such a quenching scenario will need at least eight states. Although such a description is feasible, in the following sections, only the measurements and the fits for the constructs with a quenched donor are shown. Due to the lack of observable acceptor quenching, they can be fitted with the same six state model as the polyproline data.

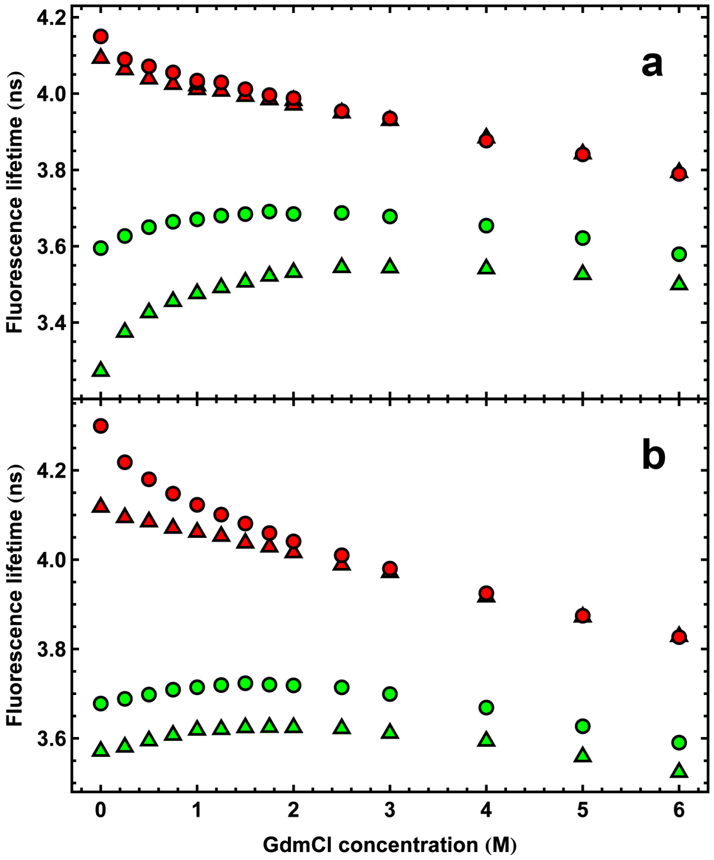


Figure 21: GdmCl concentration dependent ensemble lifetimes of the different IN variants. The values result from single exponential fits to ensemble lifetime decays measured in a magic angle configuration to avoid contributions from rotational effects. The changes in the lifetimes result from dynamic quenching between the fluorescent dye and the quenching amino acid together with an additional quenching contribution by GdmCl. The green markers show the results for Alexa 488 quenching, measured using donor-only constructs to avoid a FRET contribution. The red markers show the results from doubly labeled constructs, measured using acceptor direct excitation. Constructs containing the quenching Trp are indicated with a triangle, the control constructs lacking the Trp with a circle. **a)** IN 488W/F_594 constructs where the C11 position with the attached donor dye is quenched by Trp. Almost no difference between the acceptor lifetime of 488W_594 and 488F_594 can be seen, indicating no quenching contribution by Trp at position C60. **b)** IN 594W/F_488 constructs where the acceptor dye is at the quenched C11 position. Compared to the variant with a quenched donor, the dynamic quenching of the acceptor is less pronounced. The errors are within the size of the symbols.

Transfer Efficiency Histograms The transfer efficiency histograms of the different Integrase variants were measured the same way as for the polyprolines (See section 6.3.2). Figure 22 shows the histograms for the quenched IN 488W_594 variant, figure 23 the histograms for the Trp-free IN 488F_594 variant. As a result of the extensive purification using reverse phase chromatography, the donor-only population at zero transfer efficiency is very small for both constructs. To highlight the effect of quenching on the transfer efficiency histograms, the same instrument correction factor was used for the calculation of all the histograms. This factor was determined using free, hydrolyzed fluorescent dyes and therefore does not correct for the changed quantum yields of the dyes upon attachment to the IN constructs. To quantify the changes in the transfer efficiencies, the peaks in the histograms were fitted using a Gaussian distribution for the donor-only peak at zero transfer efficiency and using a Lognormal distribution for the peak of the FRET species at intermediate transfer efficiencies.

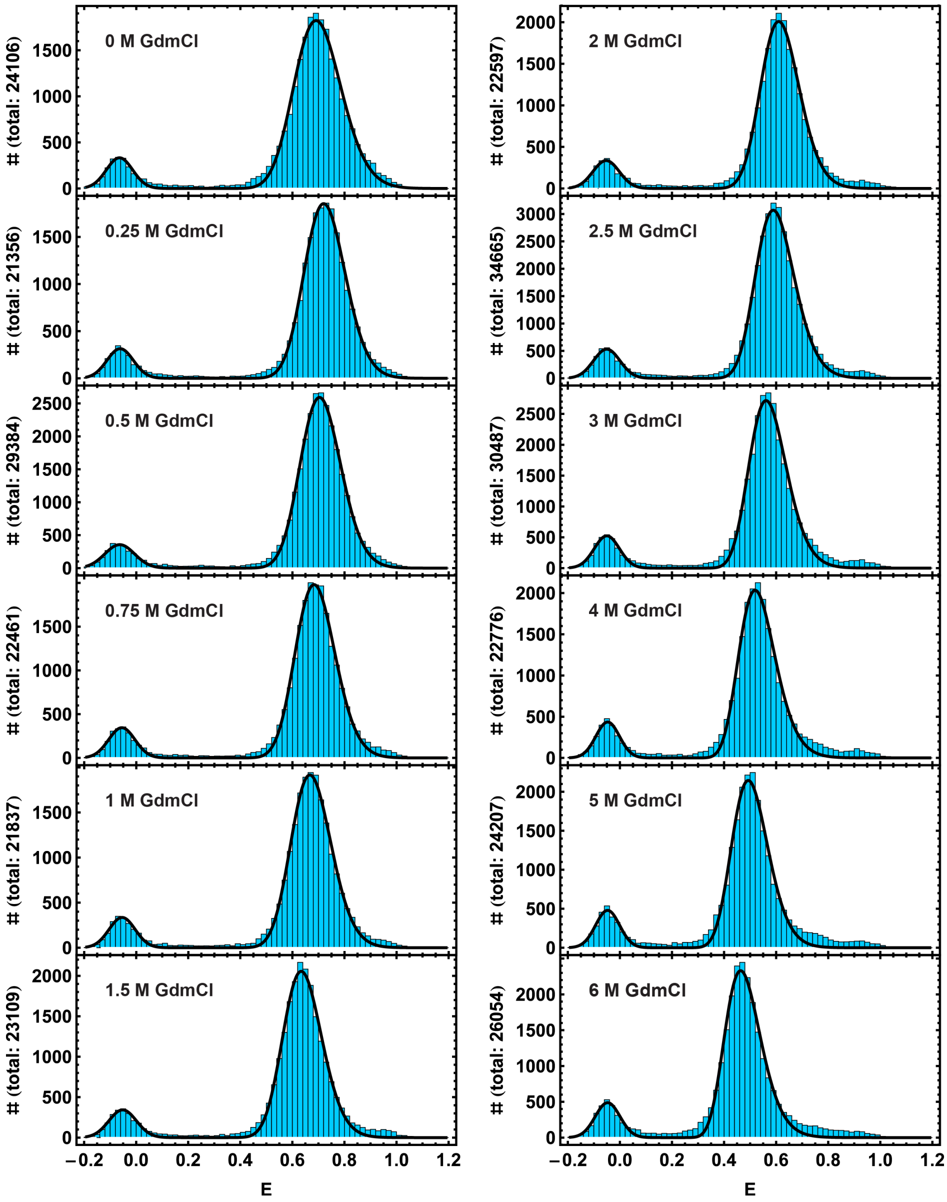


Figure 22: GdmCl dependent transfer efficiency histograms for the IN 488W_594 variant with a quenched donor. The labels indicate the corresponding GdmCl concentrations. The black line describing the histogram indicates the fit of the two populations, in which the donor-only peak, at zero transfer efficiency, was fitted with a Gaussian distribution and the observed conformational ensemble with a Lognormal distribution. The number of detected fluorescent events (bursts) is indicated on the y-axis.

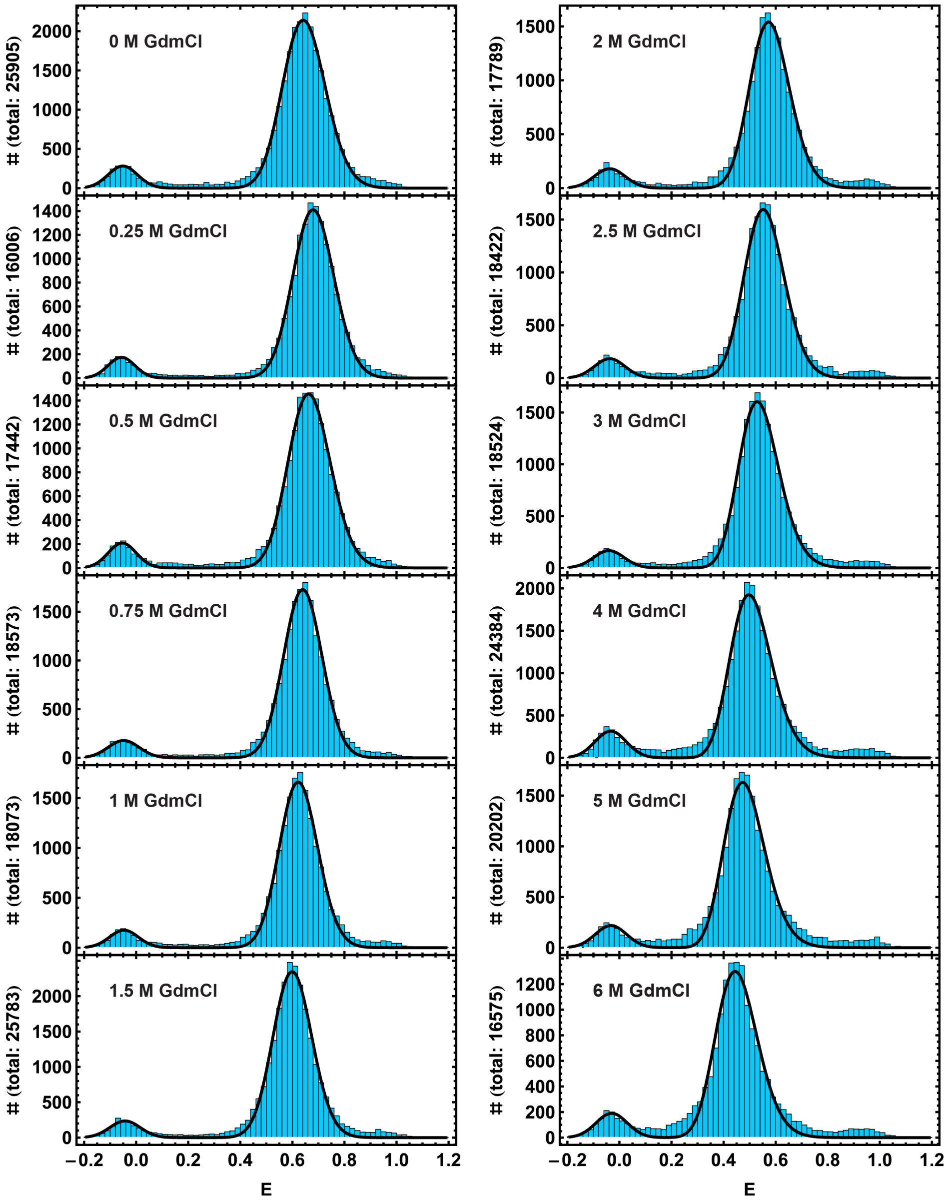


Figure 23: GdmCl dependent transfer efficiency histograms for the Trp free IN 488F_594 variant. The labels indicate the corresponding GdmCl concentration. The black line describing the histogram indicates the fit of the two populations, in which the donor-only peak, at zero transfer efficiency, was fitted with a Gaussian distribution and the observed conformational ensemble with a Lognormal distribution. The number of detected fluorescent events (bursts) is indicated on the y-axis.

The results from these fits are shown in figure 24. Similar to the results for the polyproline constructs, donor (IN 488W_594) and acceptor (IN 594W_488) quenching lead to significant changes in the observed transfer efficiencies. At zero molar GdmCl, the difference in the transfer efficiency for the two Trp containing constructs is approximately 0.13. Substitution of the Trp leads to a more converged situation with a difference in the transfer efficiencies of 0.04. This residual difference presumably results from variations in the local environment of the fluorescent dyes at the two different labeling positions leading to small changes in the corresponding quantum yields. Contributing factors may be additional quenching amino acids (See table 19) and local differences in electron densities and in hydrophobicity. A general peculiarity, found for all the investigated variants, is the increase in the transfer efficiency from 0 to 0.5 M GdmCl indicating a compaction of the unfolded protein chain. A similar behavior, found for other integrase variants, was attributed to chain compaction caused by ionic screening of the charged amino acids due to the increasing ion strength introduced by GdmCl [72]. For GdmCl concentrations above 0.5M, further increasing of the GdmCl concentration leads to a decrease of the transfer efficiencies and therefore to an expansion of the IN chain. Such an expansion is a common behavior of unfolded proteins upon increasing the denaturant concentration [39, 76]. Additionally, an increase in the GdmCl concentration results in a reduction of the quenching effects [20], leading to convergence of the observed transfer efficiencies. The transfer efficiencies of the Trp free constructs as shown in figure 24 were used to calculate the inter-dye root mean square distances used as constraints in the photon statistics fits (See section 6.4.3).

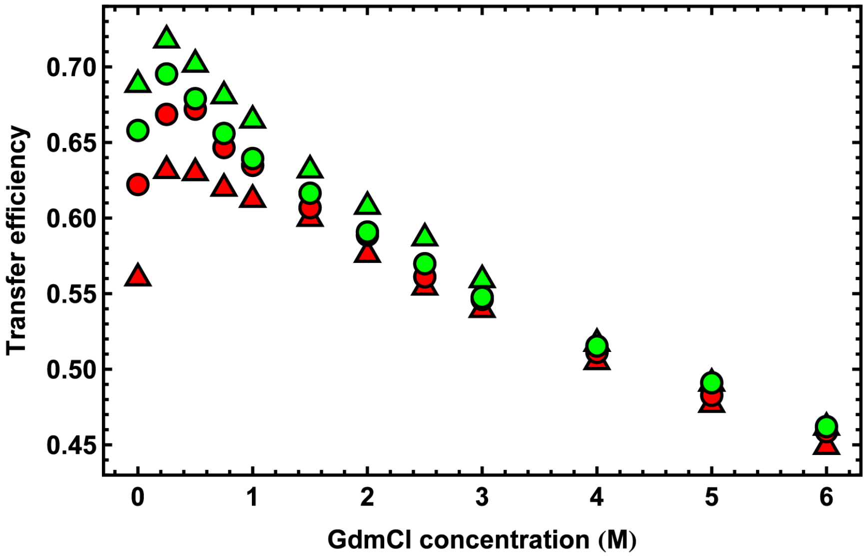


Figure 24: GdmCl concentration dependent transfer efficiencies for the different IN constructs. The transfer efficiency values represent the maximum of the Lognormal distribution fitted to the calculated transfer efficiency histograms. The green symbols represent the values for the IN 488W/F_594 and the red symbols for the IN 594W/F_488 variants. The triangles show the values for the variants with a quenching Trp and the circles for the variants where the Trp was replaced by a Phe. The values for concentrations above 4M GdmCl are equal within the error of the fits.

6.4.3 Extending the Photon Statistics Approach to Explore the Combined FRET and PET Dynamics

The procedure for fitting the GdmCl dependent ns-FCS data for the different IN variants was analogous to the procedure used for the polyprolines (See section 6.3.3). The main difference is that for the IN the donor and acceptor fluorescent dyes are not separated by a static distance. The fixed distance, in the case for the polyprolines, allowed the modelling of the FRET process using two transfer rate constants, one for the case without quenching and one for the case of transfer to a quenched acceptor dye. The diffusive dynamics of the unfolded protein chain, as found for IN, requires a model which includes a distribution of distances together with a diffusion constant to describe the interconversion rate between the different distances. To include diffusive dynamics, the rate matrix $\mathbf{K}$ has to be extended by an operator $\mathcal{L}_r$ describing the conformational dynamics as a diffusion process in a potential of mean force [28]:

$$\mathbf{K} = \mathbf{K}_0(r) + \mathbf{I}\mathcal{L}_r \tag{44}$$

I is the identity matrix, and the diffusion operator $\mathcal{L}_r$ is defined using the equilibrium inter-dye distance distribution $p_{eq}(r)$ [74, 28]:

$$\mathcal{L}_r = D\frac{\partial}{\partial r}p_{eq}(r)\frac{\partial}{\partial r}\left(p_{eq}\left(r\right)\right)^{-1} \tag{45}$$

For unfolded proteins, similar to IN, a Gaussian chain distribution was successfully applied to describe the experimental data [39]. A Gaussian chain distribution is described by the following equation [74]:

$$p_{eq}(r) = 4\pi r^2\left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2}\exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right), \tag{46}$$

where $\langle r^2\rangle$ is the mean square dye-to-dye distance. To calculate the correlation function, the operator $\mathcal{L}_r$ is discretized to a $N \times N$ tridiagonal matrix [74]. Calculating the direct product between the discretized $N \times N$ matrix for $\mathcal{L}_r$ and $\mathbf{K}_0(r)$ (of size $M \times M$) results in a block matrix $MN \times MN$ [28]. This new rate matrix is used to calculate the correlation functions in the same way as described for the polyprolines. The number of discretization steps N has to be large enough to reach a consistent result, but on the other hand small enough to still be computationally accessible. An optimization revealed an optimal value of $N = 40$. Note that the annihilation process, which reduces the double excited state D*A* by a radiation-less decay of the donor dye, was also implemented using a distance dependent decay rate $k_{SSD}(r)$. Due to the difficulty in measuring absorption spectra of the excited acceptor dye, this Förster radius was implemented as a fitting parameter.

The model extensions for including diffusive dynamics were also implemented in Wolfram Mathematica. Whereas an average fit for the polyprolines takes around 15 min, a fit including diffusive dynamics for an IN measurement can take up to 24h. Therefore implementing all the discussed optimizations (See section 6.3.3) was a crucial requirement to be able to perform such fits. As for the polyprolines, we fixed all previously known parameters. Table 3 lists the fitting parameters together with the method of determination.

Fixed Parameters	**Value and Method of Determination**
Excitation rate constant ($k_{exD}, k_{exA} = \alpha k_{exD}$)	Antibunching of Trp-free donor-only construct: 0.02 ns^{-1}, 0.001 ns^{-1}[74]
Ground state relaxation rate constant (k_D, k_A)	Ensemble lifetime measurements: 0.24-0.31 ns^{-1}
Root mean square end-to-end distances	Transfer efficiency of Trp-free construct (IN 488F_594) 4.8-6.2 nm
Relative donor-only population (r_d)	Relative histogram peak areas (donor-only / transfer population) of corresponding E histograms: 9-13%
Förster radius (R_0)	Dye emission and absorption spectra and extinction coefficients, corrected for change in refractive index and donor quantum yield: 49-54 Å
Triplet state decay times (τ_{ij})	Exponential fits of Trp-free control constructs and polynomial interpolation (IN 488F_594): 1.5 - 4.8 µs
Crosstalk (β)	Calibration measurements: 8 %
Fitting Parameters: Time origins ($t_{0_{ij}}$), Overal amplitudes (c_{ij}), Quenching rates (k_{onQ}, k_{offQ}), Inter-dye diffusion constant (D), Förster radius for donor annihilationn (R_{0SSD}), Triplet amplitudes ($c_{T_{ij}}$)	-

Table 3: Parameters for fitting the correlation functions and how they were determined. The first column (Fixed Parameters) declares the values needed by or derived using the photon statistics fit. The second column (Value and Method of Determination) shows the origin of the fixed parameters. In cases were a precise value was measured for each construct, the corresponding range is given. In the last row, the fitting parameters are shown.

Using these constraints, the ns-FSC data for the different integrase variants was fitted globally. Figure 25 shows an overview of the fit results for the integrase variant with a quenching Trp (IN 488W_594) and figure 26 for the control, lacking Trp (IN 488F_594). To avoid any bias, the same guess values were used for all of the fits. The fits include two dynamic components, first the quenching between Alexa 488 and the Trp and second the diffusive dynamics between the two fluorescent dyes. To highlight the effect of these two contributions in the ns-FCS fits, two different lines are drawn: The blue one shows the complete result of the photon statistics fit whereas the black dashed line represents the calculated component from diffusive inter-dye dynamics. Therefore, the difference between these two lines can be fully attributed to the static quenching component. For the donor quenched IN 488W_594 variant, the 0M measurement shows a significant quenching contribution in all three correlations. A hallmark for the static donor quenching process is the positive amplitude in the crosscorrelation. Since FRET dynamics between the donor and the acceptor dyes is always anticorrelated, such a positive amplitude cannot result from a dynamic FRET process. Increasing the GdmCl concentration leads to a convergence of the two lines, indicating that the observed ns-FCS is fully determined by the diffusive inter-dye dynamics and that static quenching does not contribute anymore to the overall fit. Additionally, an increase in the GdmCl concentration leads to a slight but persistent decrease in the chi-square and therefore to an increase in the fit quality. Possible explanations are additional quenching effects, not adequately modeled using a six state model, or a breakdown of the Gaussian chain assumption at low ionic strength.

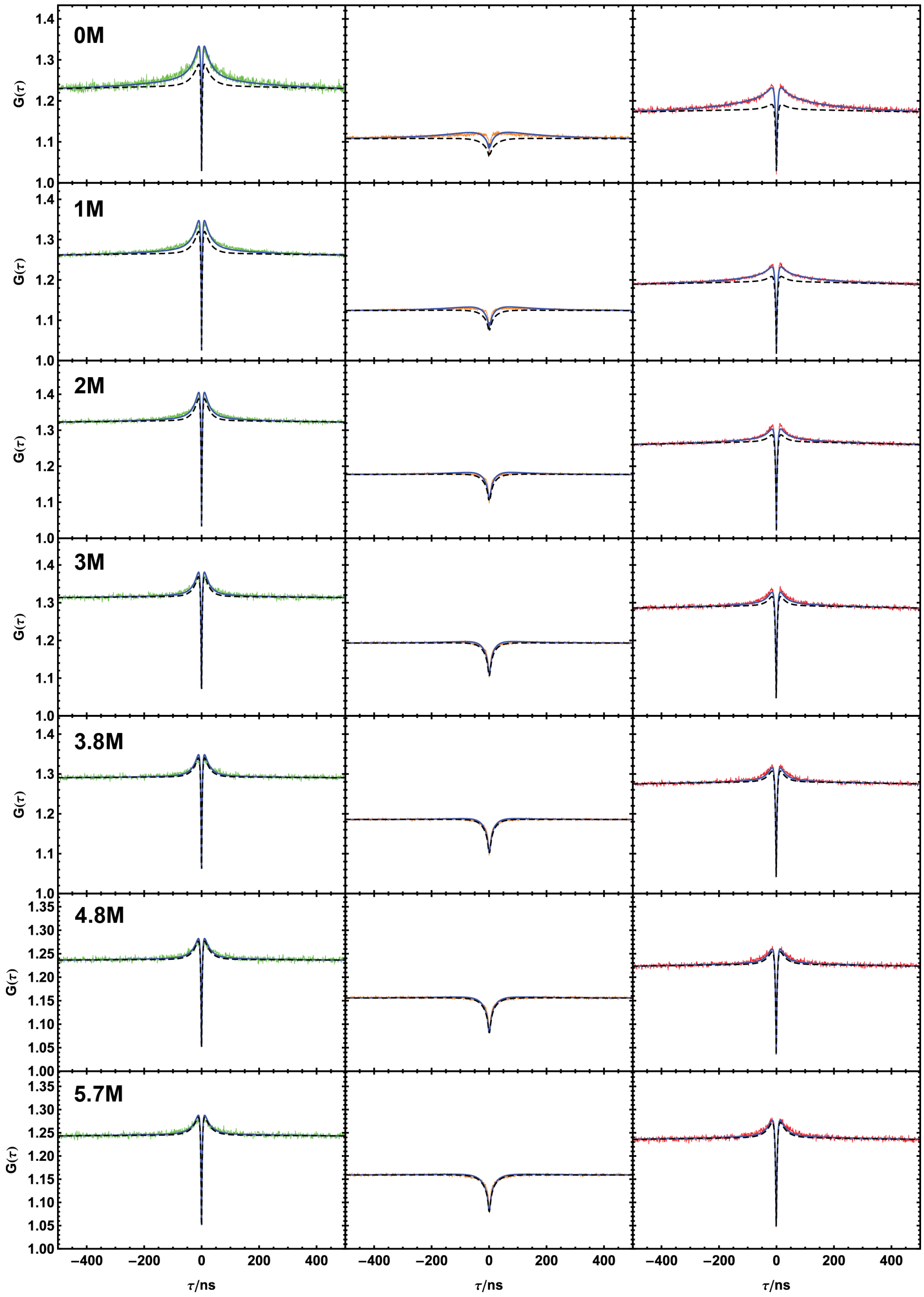


Figure 25: Representative overview of the ns-FCS measurements for the quenched IN 488W_594 variant together with the corresponding global photon statistics fits. Each row shows the donor-donor (green), donor-acceptor (orange) and the acceptor-acceptor (red) correlation resulting from the same experiment. The inset in the left column indicate the GdmCl concentration used for the measurement. The blue lines show the complete photon statistics fit including diffusive FRET dynamics and PET quenching of the donor fluorescent dye. The black dashed line shows the ns-FCS curve resulting from the diffusive dynamics between the donor and the acceptor dye (no PET quenching).

Figure 26 shows the fit results for the Trp free IN 488F_594 construct. A comparison of the 0M GdmCl ns-FCS fits of the two different variants (IN 488W_594 vs. IN 488F_594), reveals a much smaller quenching component for the Trp free construct. Although both systems were fitted with the same six state model and using the same guess values, the fitting procedure is able to distinguish between a quenched and an unquenched scenario. For the IN 488F_594 variant, adding 1M GdmCl is sufficient to remove the residual difference between the line indicating the FRET dynamics and the line for the overall ns-FCS fit. This may be an additional hint to a breakdown of the Gaussian chain approximation at low GdmCl concentrations and therefore low ionic strength. A close look at the DA/AD crosscorrelations for the IN 488F_594 reveals no observable positive amplitude, which is an additional indication of a lack of static donor quenching. The acceptor-acceptor correlation reveals a slight deviation between the experimental data and the overall ns-FCS fit. Possible explanations are a slight contribution by static acceptor quenching or an oversimplification by using a six state model for fitting the data.

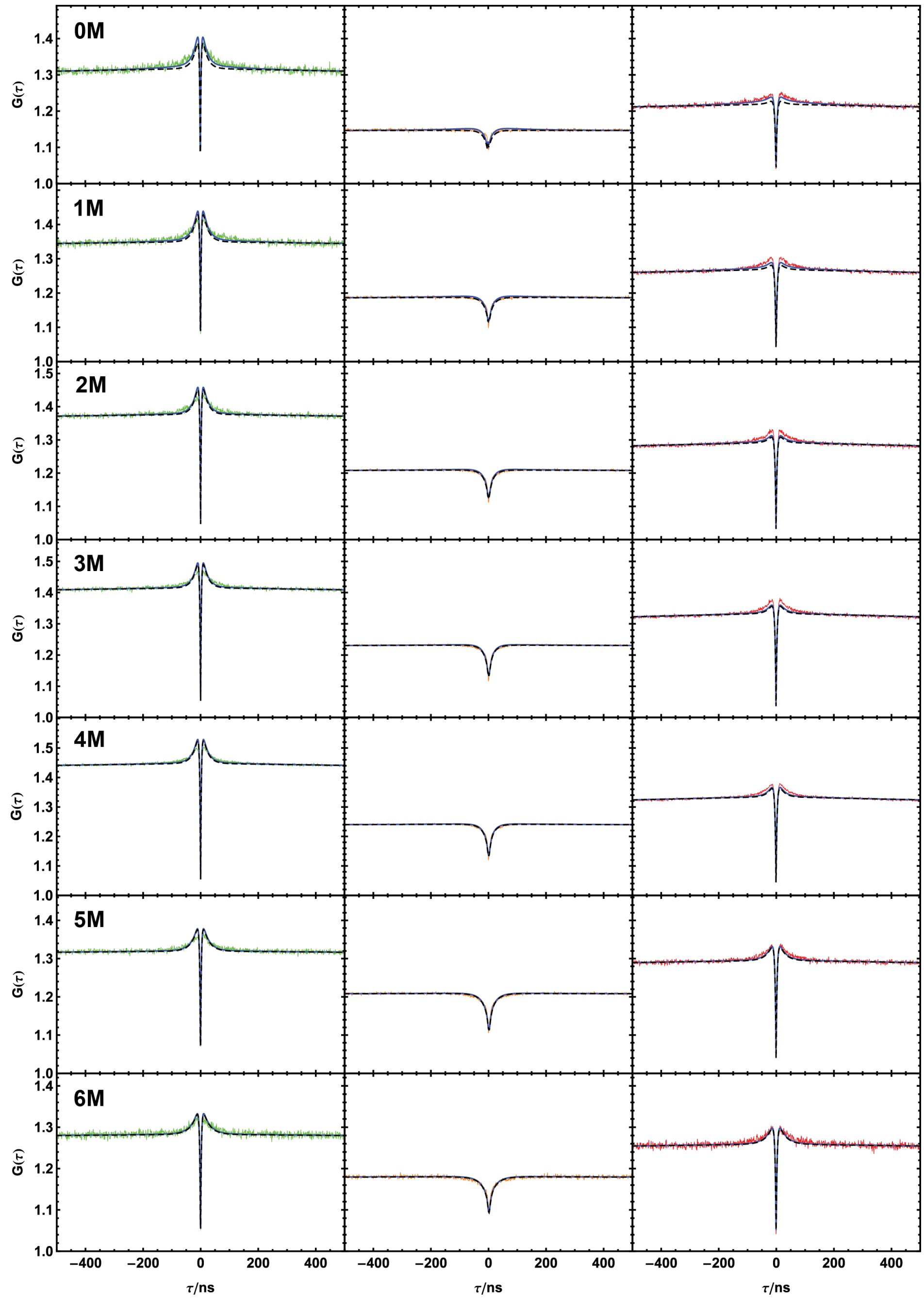


Figure 26: Representative overview of the ns-FCS measurements for the Trp free IN 488F_594 variant together with the corresponding global photon statistics fits. Each row shows the donor-donor (green), donor-acceptor (orange) and the acceptor-acceptor (red) correlation resulting from the same experiment. The inset in the left column indicate the GdmCl concentration used for the measurement. The blue lines show the complete photon statistics fit including diffusive FRET dynamics and PET quenching of the donor fluorescent dye. The black dashed line shows the ns-FCS curve resulting from the diffusive dynamics between the donor and the acceptor dye (no PET quenching).

As an additional control and to determine the quenching rate constants between the Alexa 488 donor dye on the C11 position and the quenching Trp from an independent experiment, a donor-only variant (IN 488W) was produced and ns-FCS measurements were performed. Figure 27 shows the results for the ns-FCS experiments together with a photon statistics fit using a three state model consisting of a ground state D, an excited state D* and a statically quenched state DQ. The overall quality of the fits is very good. However, the high resolution of the ns-FCS with a remarkably small noise contribution reveals a discrepancy between the fits and the experimental data, especially at low GdmCl concentrations. This may be an indication of an additional quenching process, of fast fibrational modes of the unfolded chain or of an oversimplification due to the description by a three state model. Usually, the donor-only population, which mainly consists of fluorescent particles not related to the protein construct, is quantified using the transfer efficiency histograms. In the case of the donor-only ns-FCS experiments, such quantification is not possible. Therefore the donor-only population was set to zero. Increasing the GdmCl concentration leads to a decrease of the observed ns-FCS amplitude, indicating a decrease of the static quenching effect between the Alexa 488 dye and the quenching Trp. This finding is in good agreement with the observations for the IN 488W_594 variant.

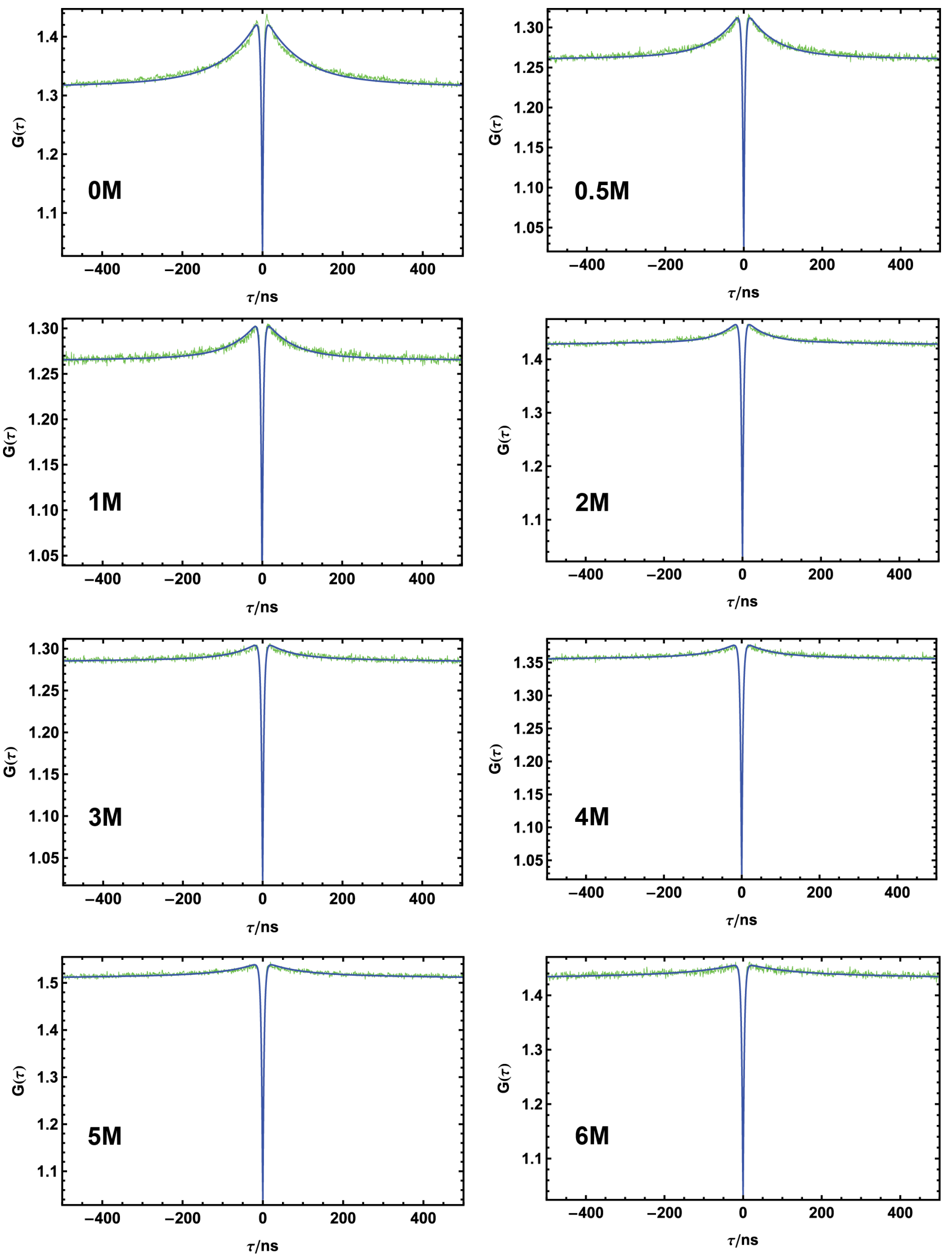


Figure 27: Donor-only control: ns-FCS measurements of the quenched IN 488W variant which carries an Alexa 488 on the C11 position but no acceptor dye on the C60 position. The data was fitted using a reduced three state photon statistics fit. The blue lines show the corresponding fit result. The GdmCl concentration is indicated in the inset.

As shown in table 3, various parameters can be derived from the photon statistics fits. The intrachain diffusion coefficient D of an unfolded protein is of particular interest since it sets a speed limit for protein folding. In the work here, it can also be used to benchmark the fitting strategy since the fitted D should primarily depend on the FRET dynamics and therefore no change should occur upon introduction of a quenching Trp. If the fitting strategy is indeed able to separate the PET and FRET dynamics, there should be no significant change in the fitted D between IN 488W_594 and IN 488F_594. Figure 28 shows the intrachain diffusion coefficients for the two variants IN 488W_594 and IN 488F_594 fitted using different strategies. The triangles show the results for the IN 488W_594 variant fitted with the 6 state model. The triangles with the black frame show the fit results obtained by fitting the quenching rates (k_{onQ}, k_{offQ}) as free parameters. The triangles with the green frame show the results obtained by constraining the quenching rates to the values obtained from the photon statistics fits of the donor-only construct (IN 488W, see figure 27). Interestingly, the difference in D between these two approaches is almost negligible which highlights the robustness of the fits using the quenching rates and the inter-dye diffusion coefficient as free parameters. The circles show the results for the IN 488F_594 variant either fitted using the six state model (closed circles) or using a four state model without the two donor quenched states (dashed circles). Upon reduction of the fitmodel to four states, the values for the IN 488F_594 variant appear to converge towards the values for the IN 488W_594 variant.

A hallmark of all of the fits is the decrease in the diffusion coefficient between 0M and 0.5M GdmCl. This effect is coupled to the compaction of IN upon increasing the ionic strength by adding up to 0.5M GdmCl (See trend of transfer efficiencies in figure 24). The increased ionic shielding due to GdmCl reduces the repulsion between the charged residues within the IN and therefore allows a compaction of the unfolded protein chain [72]. The drop in the diffusion coefficients indicates that this compaction is accompanied by a slowdown of the dynamics within the unfolded protein chain. Increasing the GdmCl concentration above 0.5M leads to an expansion of the chain together with an acceleration of the chain dynamics. Independent of the variant and of the fitting strategy used, above 4M GdmCl the diffusion coefficients converge. In terms of the obtained diffusion coefficients, a very similar range and an increase with increasing GdmCl was found for a completely different system, namely for the small cold shock protein Csp *Tm* [39].

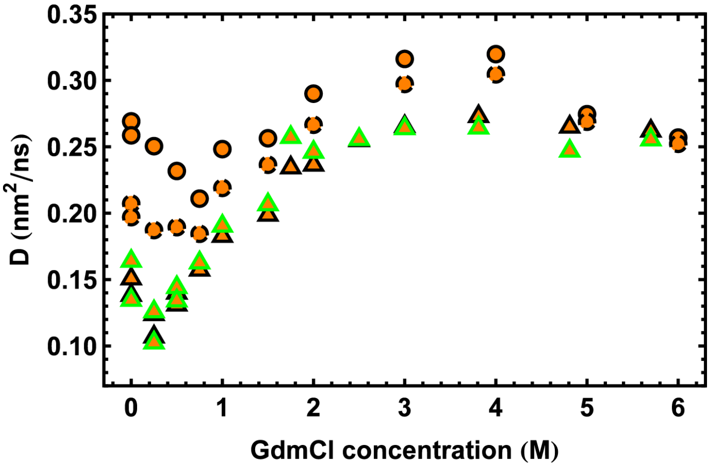


Figure 28: GdmCl dependent inter-dye diffusion coefficients obtained by global ns-FCS fits. IN 488W_594 (triangles) fitted with a six state model with the quenching rates (k_{onQ}, k_{offQ}) and the diffusion coefficient (D) as free parameters (black) or the quenching rates fixed to values obtained from the donor-only control (green) (Fits shown in figure 25). The Trp free variant, IN 488F_594 (circles), was fitted with either a six state model with free fitting parameters (solid black) or a four state model without quenched states (broken) (Fits shown in figure 26).

The rates for the formation (on-rate) and dissociation (off-rate) of the statically quenched complex between the Alexa 488 donor dye and the Trp are interesting parameters obtained from the photon statistics fits. A comparison of the quenching on-rates to the intrachain diffusion coefficients D can lead to novel insights about the connection between short-range contact dynamics in relation to the timescales of the global reconfiguration time of the protein chain. Figure 29 shows the result for the on-rates (a) and the off-rates (b) of the different constructs. The squares report on the rates obtained by fitting the donor-only constructs (IN 488W) using a three state model (See figure 27). The triangles show the rates for the IN 488W_594 and the circles for the Trp free IN 488F_594 variants, both fitted using a six state model. A common behavior for all three constructs is a decrease in the on-rate with increasing GdmCl concentration. This may result from the expansion of the unfolded chain or from a reduced probability of forming a statically quenched complex upon increase of GdmCl concentration. Interestingly, compared to the derived intrachain diffusion coefficients, there is no sign of any special behavior of the on-rates between 0M and 0.5M GdmCl. A comparison of the on-rates for the different variants reveals the highest values for the donor-only construct (IN 488W) fitted with a three state model. The introduction of the acceptor dye Alexa 594 together with a global fit using a six state model leads to the values for IN 488W_594 represented as triangles. There is no obvious reason for the difference in the rates for these two constructs since the static donor quenching should not be affected by the introduction of an acceptor dye. Most probably, this difference emerges from the different fitting strategies. Using identical fitting strategies and guess values, a comparison of the on-rates between IN 488W_594 and IN 488F_594 can be used to derive more significant insights. Here, the removal of the Trp leads to a four-fold drop of the observed on-rates for the quenching process. The residual quenching rate, observed for the IN 488F_594 variant, results from a combination of small quenching contributions by amino acids other than Trp together with an overlap of the FRET signal by the PET quenching process due to the fit which still includes static quenching. Increasing the GdmCl concentration leads to an increase in the observed off-rates. This indicates a destabilization of the quenched complex either by directly affecting the Alexa 488-Trp interaction or by expanding the protein chain and therefore introducing the possibility of an increased strain on the complex. A similar reduction of the complex stability between the fluorescent dye MR121 and a quenching Trp upon increase of the GdmCl concentration was found by Dose *et al.* [20]. The observed differences in the off-rates between the three constructs are slightly more difficult to explain. The off-rate for the Trp-free variant IN 488F_594 shows almost no GdmCl dependence, indicating the fitting of a residual effect in the ns-FCS not directly related to quenching of the Alexa 488 dye. The difference between IN 488W_594 and the corresponding control IN 488W most probably arises again from the different fitting strategies. These results indicate that an exact assignment of the correct on and off-rates for the formation of the statically quenched complex is difficult to achieve without further constraints from control experiments. The minor changes in the intrachain diffusion coefficients D upon constraining the quenching rates to the ones obtained from the donor-only control (IN 488W), as shown by the triangles with the green frame (See figure 28) highlight this situation. An additional important aspect, leading to a better understanding of the fitting behavior, is a comparison of the equilibrium constants for the quenched complexes.

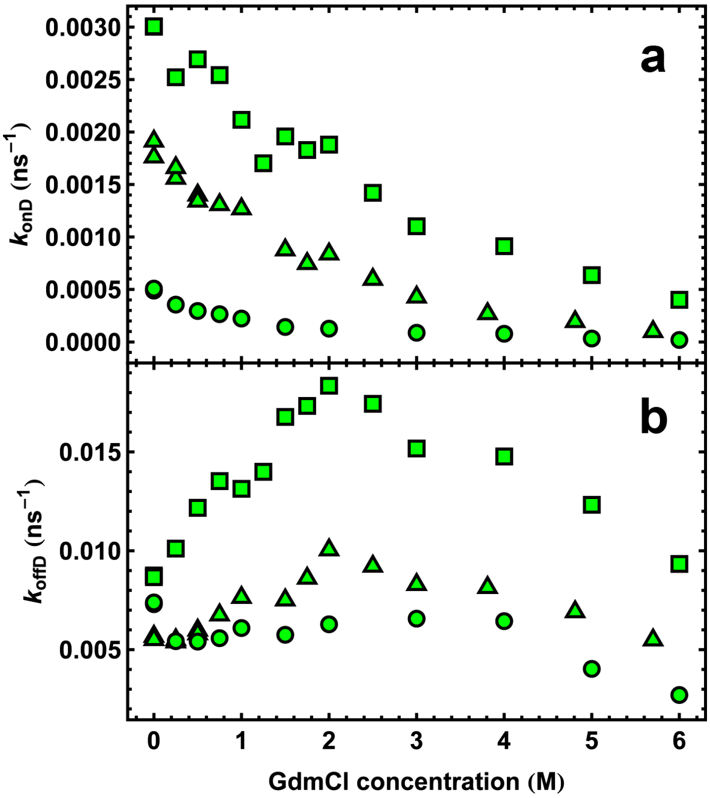


Figure 29: GdmCl dependent PET quenching rates of various IN variants. Squares: Donor-only IN 488W variant fitted with a three state model (Fits shown in figure 27). Triangles: IN 488W_594 variant globally fitted with a 6 state model (Fits shown in figure 25). Circles: Trp free IN 488F_594 variant globally fitted with a 6 state model (Fits shown in figure 26). **a)** On-rates for donor quenching. **b)** Off-rates for donor quenching.

The on- and off-rates obtained by the photon statistics fits allow the calculation of the equilibrium constants for the statically quenched complex. Figure 30 shows the results obtained from the rates shown in figure 29. This representation reveals that the assigned amount of quenching for the IN 488W_594 and the corresponding control IN 488W are identical within the experimental error (squares & triangles). For the Trp free IN 488F_594 variant, the equilibrium concentration of quenched Alexa 488 is five times smaller. This indicates that the equilibrium quenching constant is a more robust parameter than the rates of the quenching process. Figure 30 also shows a remarkable drop in the equilibrium constant between 0M and 2M GdmCl. For concentrations larger than 2M GdmCl, the equilibrium population of the quenched complex is below 10% for all of the variants. Such small equilibrium populations lead to very small amplitudes in the ns-FCS curves and therefore a large uncertainty in the photon statistics fits. Most probably this is the reason for the observed discontinuities in the off-rates above 2M of GdmCl (See figure 29(b)). Therefore, in order to obtain reliable results from the quenching rates derived by a photon statistics fit, the statically quenched population should be larger than 10%.

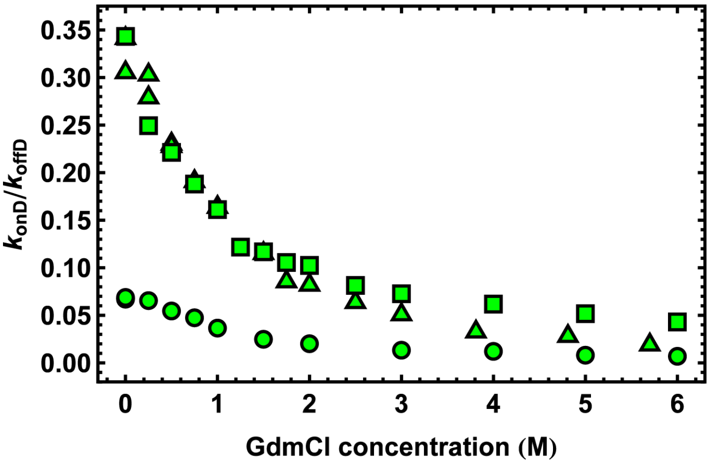


Figure 30: GdmCl dependent equilibrium quenching constants, calculated from the on- and off-rates obtained from the photon statistics fits. Squares: Donor-only IN 488W variant fitted with a three state model (Fits shown in figure 27). Triangles: IN 488W_594 variant globally fitted with a 6 state model (Fits shown in figure 25). Circles: Trp free IN 488F_594 variant globally fitted with a 6 state model (Fits shown in figure 26).

Finally, the photon-statistics fits allow novel insights into the very fast processes of the ns-FCS curves. From the so called antibunching region, which spans the first few ns in the ns-FCS, information about the radiative rates, the FRET rate and the behavior of the double-excited state D*A* can be derived. To model the experimentally found behavior, especially the amplitude of the antibunching peak in the crosscorrelation, a non-radiative decay process of the double-excited state D*A* has to be included. To date there is no characterization of this so called singlet-singlet annihilation process between the two excited dyes Alexa 488 and Alexa 594 available. Simulating ns-FCS curves revealed that the two extreme mechanisms, either transfer from Alexa 488* to Alexa 594* or the other way round, result in almost no observable differences for the experimental conditions used here. Therefore it was implemented in the photon statistics fit similar to the FRET process with a non-radiative transfer from the donor fluorescent dye Alexa 488* to the acceptor Alexa 594*. It was assumed that the annihilation process has identical distance dependence as the FRET process. It was implemented using a distance dependent annihilation rate constant based on an annihilation radius similar to the Förster radius. The so-called annihilation radius was a fitting parameter in all of the fits. Figure 31 shows the fit results for this annihilation radius. The triangles represent the values for the IN 488W_594 and the circles for the IN 488F_594. The squares, for comparison, show the fixed fitting values for the Förster radius of the FRET process calculated from the change in the refractive index of the solvent. The first remarkable result is the small scatter in the obtained values for the annihilation radii, indicating that the photon statistics fits are able to extract this parameter with a small uncertainty. Introducing a quenching Trp should not change the behaviour of the annihilation process. Comparing the values for the two different IN variants indeed shows no significant difference between the one with (triangles) and the one without Trp (circles). In comparison to the Förster radii of the FRET process, the biggest difference is the reduction of around 1 nm for the annihilation radii. By considering this process in a manner similar to the Förster process, the only parameters that can change are the absorption spectrum of the acceptor dye and/or the corresponding extinction coefficient (See equation 3). A reduction in the annihilation radius indicates a smaller overlap integral J. Possibly this results from a redshift of the acceptor absorption in the excited state. The fitting results in figure 31 show a small turnover effect between 0M and 0.8M GdmCl. The origin of this effect is not clear as GdmCl should only change the refractive index and therefore a linear dependence for the annihilation radius is expected. Possible explanations are a fitting artifact or an additional effect contributing to the fast antibunching timescales as observed in the cross correlation ns-FCS measurements.

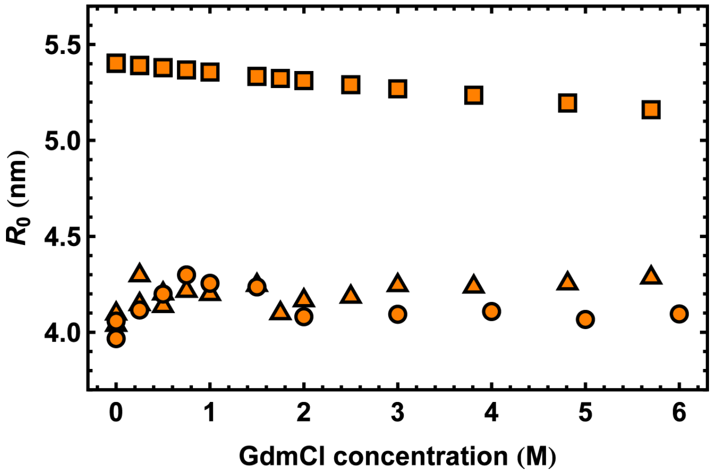


Figure 31: GdmCl dependent Förster and singlet-singlet annihilation radii. Squares: Förster radii for the S1->S0 FRET transfer used as fixed parameters in the photon statistics fits. Triangles and Circles: Annihilation radii for the S1-S1 singlet-singlet annihilation used as a free parameter in the photon statistics fits. The triangles show the values for IN 488W_594 variant and the circles for the Trp free IN 488F_594 variant.

The experimental results together with the photon statistics fitting approach show for the first time the possibility to separate and quantify FRET and PET quenching in the same molecule at the same time. The N-terminal domain of HIV-1 integrase, an intrinsically disordered peptide, was successfully used to study the short-range quenching dynamics between the Alexa 488 fluorescent dye and a single intrinsic Trp residue. As a reporter for long-range protein chain dynamics, Alexa 594 was introduced as an acceptor fluorescent dye allowing to study FRET dynamics. The quality of the fits and therefore the obtained rates of the various involved processes together with the rigorous controls highlight the robustness of this general approach.

6.4.4 Mechanistic Considerations

A comparison between the intrachain diffusion coefficients D and the on-rates of the static quenching process can lead to novel insights about the connection between local and global dynamics in an unfolded protein chain. The photon statistics fit reveals the rate of successful formed static complexes between the Alexa 488 fluorescent dye and the quenching Trp. As shown for the free dyes, this rate can be different from the diffusion-limited rate since the formation of the complex can have an activation barrier. In contrast, dynamic quenching is thought of as a diffusion-limited process were each encounter between Alexa 488 and the quencher within a certain distance leads to a successful quenching event (See table 1). A comparison of the fluorescence lifetimes for the quenched and the unquenched constructs to the rates of static complex formation can yield the probability of static complex formation upon close encounters of the dye and the quencher. Steric effects such as the ones found in a protein chain can alter this probability with respect to the values found for just free dye and quencher in solution. The ensemble lifetimes and ns-FCS measurements of the donor-only constructs IN 488W and IN 488F at various GdmCl concentrations were used to calculate these probabilities. Figure 32 shows the results with a polynomial fit which was then used for correcting the on-rates obtained by the photon statistics fits. A remarkable effect is the decrease in the probability of forming a statically quenched complex in comparison to free dye and Trp in solution. Where the free dye and Trp need approximately 3 encounters (See table 1) until a quenched complex is formed, they need at least 6-10 encounters when attached to the integrase chain. Increasing the GdmCl concentration above 3M seems to further decrease this probability. The values from the fit, indicated as a blue line in figure 32, were used to correct the on-rates for the IN 488W_594 variant, obtained by the photon statistics fits. The results of this correction are shown as green squares in figure 33.

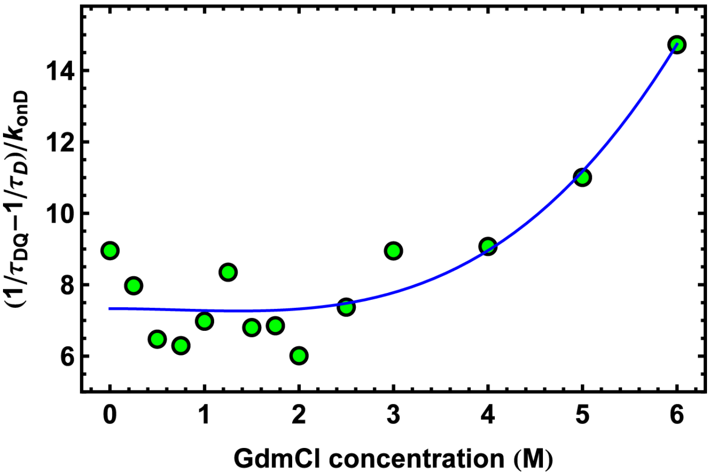


Figure 32: GdmCl concentration dependent number of collisional quenching events needed until a statically quenched complex is formed. The values result from measurements of the quenched donor-only variant (IN 488W). A comparison of the dynamic quenching rates obtained from ensemble lifetime measurements (τ_D, τ_{DQ}) to the on-rates (k_{onD}) from the photon statistics fits allows the calculation of these values. The blue line shows a polynomial fit to the data.

A direct comparison of these on-rates to the intrachain diffusion coefficients D is not intuitive. Therefore the mean first-passage time theory of Szabo, Schulten and Schulten was used to convert the obtained diffusion coefficients, together with the knowledge of the mean squared distance between the dyes from the transfer efficiency histograms, to an association rate of the chain ends [108]. Assuming Gaussian chain behavior, the following equation can be used to approximate the association rate [61]:

$$k_{ass} = \frac{4\pi D\alpha}{(2\pi\langle r^2\rangle/3)^{3/2}} \tag{47}$$

where D is the intrachain diffusion coefficient and α is the interaction distance at which quenching occurs. For calculating the association rate constants, the mean squared distances derived from the histograms of the Trp free variant IN 488F_594 were used to avoid any significant quenching contribution in this value. Due to the linear scaling of the mean squared end-to-end distance of a Gaussian chain with the number of amino acid residues, the mean squared inter-dye distance was rescaled to the distance between the Alexa 488 dye and the quenching Trp by a factor of $\frac{12}{49}$ where the numbers indicate the according sequence separations. The exact value for α is not known. Therefore, as discussed in section 6.2, a value of 0.5 nm was used here. The resulting association rates are shown in figure 33.

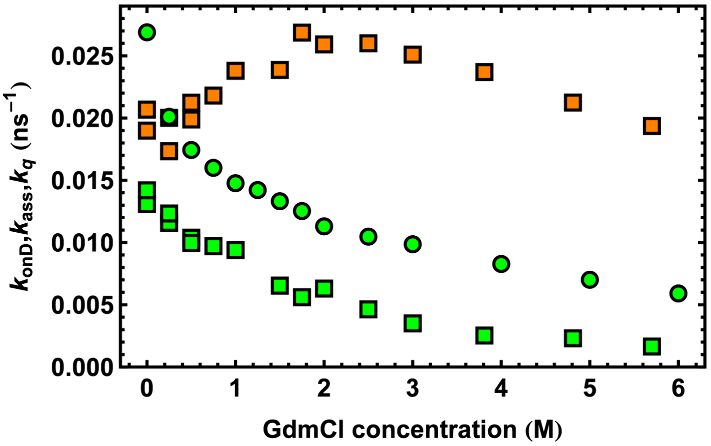


Figure 33: Comparison between the diffusion-limited on-rates from the photon statistics fits and from the fluorescence lifetime experiments to the association rates calculated using the Szabo, Schulten and Schulten mean first-passage time theory. Green squares: Static quenching on-rates from the photon statistics fits of the IN 488W_594 variant multiplied by the corresponding correction factor (See figure 32) to get the diffusion-limited quenching constants. Orange squares: Association rates calculated from the mean squared dye-to-dye distance together with the inter-dye diffusion coefficient from the photon statistics fits of the IN 488W_594 variant assuming a Gaussian chain behavior. Green circles: Dynamic quenching rates obtained from the fluorescent lifetime data of the donor-only constructs IN 488W and IN 488F

A remarkable finding is the good agreement of the on-rates at low GdmCl concentrations. This shows that the two orthogonal methods, namely quenching (green symbols) and FRET dynamics (orange symbols) can lead to very similar results in terms of the observed chain dynamics. This is also an indication that the fit does perform remarkably well in separating the quenching on-rates and the FRET dynamics in the combined ns-FCS. An increase in the GdmCl concentration leads to a divergence of the rates. Whereas the local PET quenching rate decreases, the association rate, calculated from the global intrachain diffusion coefficient, increases. This indicates that the increase in overall chain dynamics is not coupled to an increase in local contact formation rates between the involved residues. To show the reliability of these static quenching results, the dynamic quenching rates, obtained from ensemble fluorescence lifetime experiments are shown as green circles. The overall trend is very similar to the one found for the static quenching. However, there is a systematic offset in the found quenching rates for the two different methods. There are two possible explanations for the difference between the FRET dynamics and the PET quenching behavior: Either GdmCl interacts with the Trp residues and/or the Alexa 488 fluorescent dye in a way that dynamic and static quenching is reduced, or the concept that an increase in global protein chain dynamics is coupled to an increase in local dynamics is wrong. Interaction of GdmCl and unfolded proteins has been postulated to occur via binding of GdmCl to the amino acids [67, 83]. If such a binding introduces a local stiffness in the chain, short-range and long-range vibrational modes may be affected differently. Local stiffness reduces the probability of complex formation between the fluorescent dye and the Trp, and therefore the quenching rate will decrease. However, on a global scale, the introduction of chain stiffness may reduce the number of accessible conformations and therefore decrease the reconfiguration time, leading to an increase in the observed intrachain diffusion coefficient. Such behavior could explain the observed effects. Recent experiments, performed on a slightly different IN variant, showed a decrease in internal friction with increasing concentrations of denaturant [101]. This finding is contradictory to the above postulate of increasing stiffness due to binding of GdmCl to the unfolded peptide chain. For both reporter systems, a decrease in the rates is found above 3M

GdmCl. This is most probably the effect of increasing solvent viscosity at increased GdmCl concentrations. This short example shows how the combined approach using FRET and PET quenching, on the same molecule at the same time, can lead to novel insights about protein dynamics and interactions. These proof of principle experiments should pave the way for more advanced experiments addressing the most controversial topics in today's protein folding and dynamics research. Highlighting the contributing effects and showing reliable strategies to quantitatively describe them should sharpen the general attention to the obstacles introduced by performing such single-molecule experiments.

6.4.5 Robustness of Photon Statistics Fits

Developing complex fitting strategies, such as the photon statistics fits presented in this work, always needs careful testing of the overall robustness. To check if the fits are able to locate global minima, synthetic test data with known values should always be fitted. We generated various datasets with a random variation of parameters spanning the complete range as found in the experimental data. After adding noise levels similar to the experimental ones, the identical fitting strategy as for the integrase data was applied. For the constrained parameters, the input values for generating the synthetic data were used. First the triplet components were fitted using an exponential decay. The triplet decay times were fixed and the ns-FCS data was fitted globally using the photon statistics fitting approach. Figure 34 shows a representative unbiased selection of such generated data together with the photon statistics fits in black.

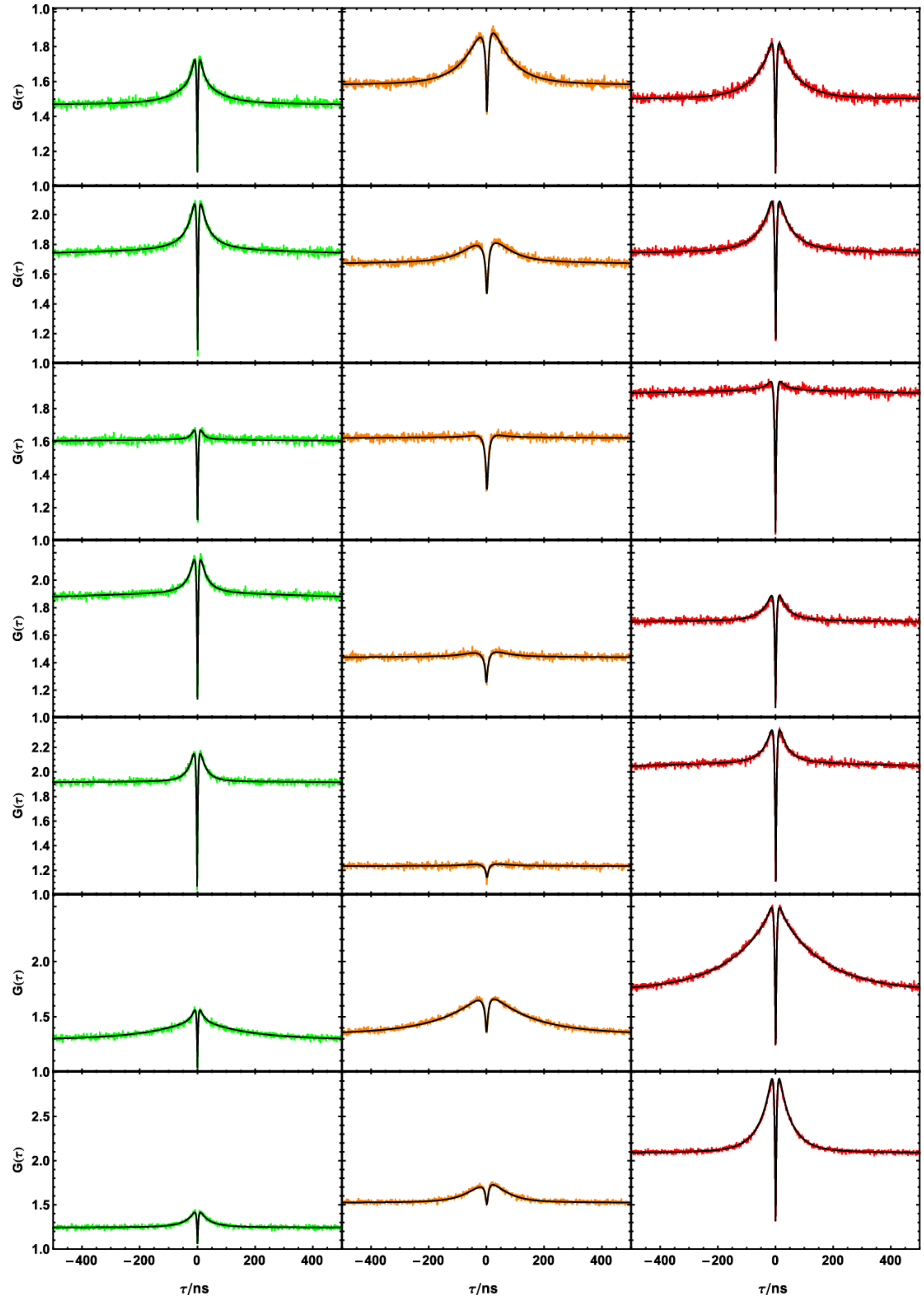


Figure 34: Representative overview of simulated ns-FCS data together with the corresponding global photon statistics fits in black. Each row shows the donor-donor (green), donor-acceptor (orange) and the acceptor-acceptor (red) correlation resulting from the same parameter set.

Since the mean squared dye-to-dye distance is derived from fitting the single-molecule histograms, the biggest error is introduced into the photon statistics fits by this fixed parameter. To acquire a more realistic picture of the fitting performance, this value was constrained within a range of +/- 5% of the value used for the simulation of the ns-FCS data. Figure 35 shows the identity plots between the values used to generate the synthetic data and the results obtained by the photon statistics fits for the most important photon statistics parameters. The error bars result from the introduced error in the fixed mean squared dye-to-dye distance. For most of the cases, the fit results are very close to the input values, indicating a robust fitting procedure.

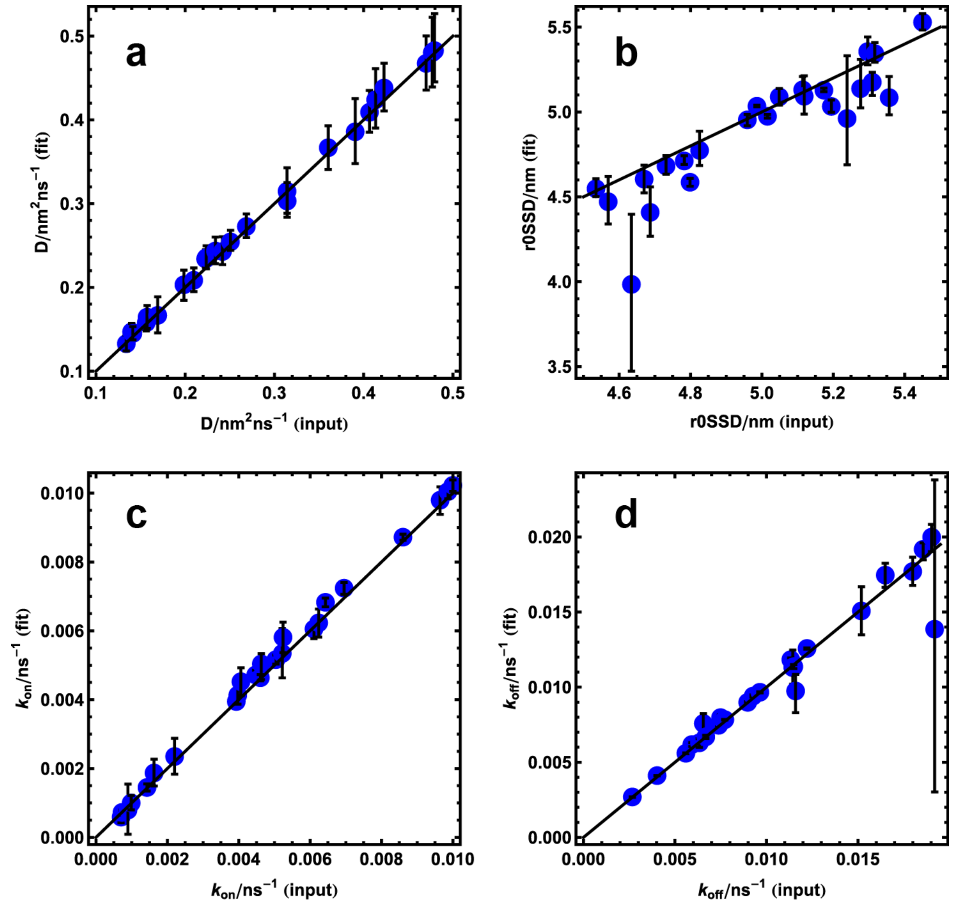


Figure 35: Identity plots for the fitting of the simulated ns-FCS data. Here the most important parameters obtained by the photon statistics fits are shown. On the x-axis the input values for the simulation of the ns-FCS data are indicated. The y-axis shows the results from the photon statistics fits. The identity line is drawn in black. The error bars result from a +/-5% variation in the mean squared dye-to-dye distance introduced as a fixed fitting parameter. **a)** Inter-dye diffusion coefficients D. **b)** Singlet-singlet annihilation radii. **c)** On-rate for the static quenching process of the donor dye. **d)** Off-rate for the static quenching process of the donor dye.

Figure 36 shows the fit result for the triplet parameters. Here the deviation between the input parameters and the fit results from the exponential fits of the triplet states are much more pronounced. However, due to the separation in the timescales between the triplet decay times and the other photon statistics effects such as the PET quenching or the FRET dynamics, this uncertainty in the triplet states does not have a significant impact on D and on the quenching rates (See figure 35).

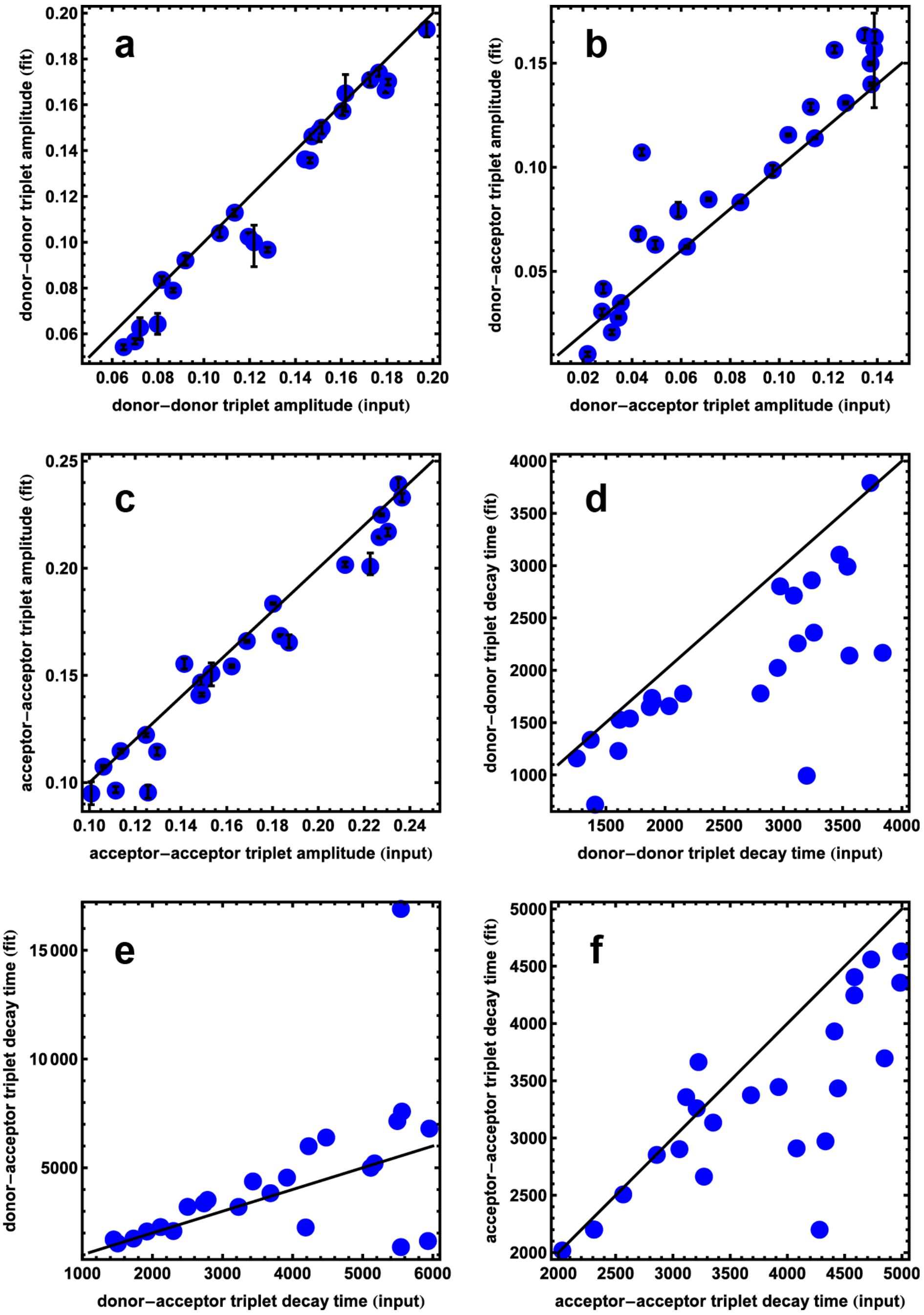


Figure 36: Identity plots for the fitting of the simulated ns-FCS data. Here the triplet parameters, fitted using exponential decays, are shown. On the x-axis the input values for the simulation of the ns-FCS data are indicated. The y-axis shows the results from the photon statistics fits. The identity line is drawn in black. The error bars result from a +/- 5% variation in the mean squared dye-to-dye distance introduced as a fixed fitting parameter. **a)** Donor-donor triplet amplitude. **b)** Donor-Acceptor triplet amplitude. **c)** Acceptor-acceptor triplet amplitude. **d)** Donor-donor triplet decay time. **e)** Donor-acceptor triplet decay time. **f)** Acceptor-acceptor triplet decay time.

The fits of the synthetic data show that the fitting strategy is indeed able to find the global minimum. However, care should be taken in fitting experimental data, and the robustness of the fits should be validated. A variation of the guess values, for example, can already indicate if the fits converge to the same minimum or not. In the case for the integrase data, the imprecision in the fitted triplet parameters does not pose a significant problem due to the clear separation of the involved timescales. Using such a methodology for alternative experimental systems requires a careful check if this requirement is still satisfied. If not, alternative fitting strategies have to be applied which are able to fit the triplet states with a more reliable outcome.

7 Conclusions and Outlook

Understanding the conformations and the dynamics of unfolded proteins is a prerequisite for a mechanistic understanding of protein folding. Novel insights into these processes will have a fundamental impact in a variety of related research fields such as molecular dynamics force field development. Single-molecule methods, as the ones used in this work, are of particular interest since they allow studying the conformations and the dynamics of different species in an equilibrium system. This work consists of three main parts, covering different aspects of protein folding and dynamics together with the development of novel experimental tools to for single-molecule spectroscopy. In the first part, experimental tools for single-molecule FRET experiments were developed and general limitations such as confinement of the fluorophores and the quenching of fluorescent dyes by natural amino acids were investigated. In the second part, the temperature-induced collapse of unfolded proteins and protein chaperone interactions were investigated. In the last part the obtained knowledge was used to develop a new approach to measure short- and long-range dynamics within a single molecule at the very same time.

In the last years, more and more evidence emerged that protein folding cannot be described as an isolated behavior of a single-molecule in solution but embedding such a molecular system into a cellular environment can lead to a different picture. It was found that factors such as molecular crowding, cotranslational folding and molecular chaperones can have an impact on protein folding and dynamics [26]. In this work, we systematically investigated interactions between unfolded rhodanese and the GroEL chaperone. It was shown that rhodanese is bound to GroEL as a partially structured folding intermediate with a rather well defined ensemble of conformations. Next the folding kinetics of rhodanese, encapsulated within GroEL/ES was studied. Depending on the used rhodanese variants, different effects on the folding kinetics were found. This indicates that there is no universal mechanism of GroEL/ES action and the specific mechanism may be different from protein to protein. These results leave the "Anfinsen folding cage" as the only model partially describing the function of the GroEL/ES chaperone system [42]. In addition to this important finding, these experiments showed for the first time the possibility to obtain information about protein conformations using single-molecule FRET in the presence of restricted chromophore flexibility. The GroEL experiments also revealed fundamental limitations in performing single-molecule FRET experiments. Therefore surface immobilization methods and photoprotection strategies were developed to solve these limitations. The improvements in fluorophore stability also allowed the measurement of the temperature-induced collapse of unfolded proteins. Here, a compaction due to an increase of intramolecular hydrogen bonds at increasing temperatures was found. The short distances together with the reduced mobility of the fluorescent dyes within the GroEL/ES chaperone limited the applicability of FRET as a reporter system for protein chain dynamics. Therefore a novel strategy to combine short-range (PET) and long-range (FRET) reporters on the same molecule was developed. Fluorescently labeled polyprolines carrying an additional Trp as a PET quencher were produced to study the effect of PET quenching on a static FRET system. A photon statistics fit approach was developed to globally fit the resulting ns-FCS data. To combine FRET and PET dynamics within a single molecule, new variants of the N-terminal domain of HIV-1 integrase, an intrinsically disordered protein, were developed. Site specific labeling allowed either donor or acceptor quenching by an intrinsic Trp residue. Extending the photon statistics fitting approach to include FRET dynamics allowed the separation of FRET and PET quenching dynamics by fitting the ns-FCS results globally. A comparison of the obtained rates revealed a different behavior for short-range and long-range dynamics upon increase of the GdmCl concentration. Whereas the global inter-dye diffusion constant increases with increasing GdmCl concentration, the local PET on-rate decreases, indicating a slowdown of the quenching dynamics. This points to a binding of GdmCl to the unfolded protein chain and therefore to an increase in the persistence length. It was also shown how the rigorous quantification of all the involved effects can be used to correct single-molecule

FRET data and therefore to obtain more reliable results from these experiments.

All the FRET experiments, performed for this work, to directly measure unfolded protein chain dynamics within the GroEL/ES chaperone system resulted in no observable effect besides the rotational components resulting from the large GroEL(/ES)-substrate complexes. A combination of a short- and long-range reporter systems, as the one introduced here, has the potential to lead to new insights about how the GroEL/ES chaperone does affect the chain dynamics of an encapsulated substrate protein. Future directions in the development of the photon statistics fitting approach will be the inclusion of triplet processes into the kinetic description and a global fit of all the available timescales (ns-ms) and therefore including slow dynamics and intermolecular diffusion processes. A more refined description of the singlet-singlet annihilation processes will lead to improved fits for the fast antibunching timescales in the ns-FCS curves. To use the correlated FRET and PET information, new data analysis methods have to be developed. Possible candidates are likelihood methods and multidimensional FCS analysis. To describe the experimental situation for more complicated systems, rotational effects and dye-dye quenching can be included. Separating the quencher and the fluorescent dyes on different but interacting molecular systems will lead to novel insights about the coupling of dynamics and intermolecular interactions.

8 Appendix

8.1 Surface Immobilization Experiments

To be able to perform robust and reproducible surface experiments, we developed a flow-cell platform as shown in figure 37 and a software to automatically identify and measure single molecules on the surface shown in figure 39. Due to a lack of documentation by the manufacturer, we used reverse engineering methods to control the piezo stage and the confocal microscope. The communication between the controlling software[20] and the piezo stage was recorded using a serial port monitor[21]. Since the communication to the confocal instrument which controls the shutters and the photodiode is implemented using a CAN bus and therefore a dedicated hardware interface we had to use a special monitoring tool by the supplier of the CAN hardware[22]. Complete control of the hardware allowed us to remotely control the laser shutter and therefore to reduce the exposure time and photobleaching. Since sample application and liquid handling of the flow cell is performed by standard tubing and connections, buffer exchange and even kinetic experiments using a syringe pump can be performed routinely. To immobilize the protein of interest, various methods can be applied, for example, immobilization on biotinylated polyethylene glycol (PEG) surfaces or on biotinylated lipid bilayers. The protein of interest can either be biotinylated and directly linked to the surface *via* streptavidin or enclosed in biotin-containing lipid vesicles, which are subsequently bound to the surface *via* streptavidin. For the functionalized surfaces either ionic immobilization of PLL-PEG-Biotin or cover glasses with covalently attached PEG-biotin can be used.

An additional advantage of the flow-cell hardware shown in figure 37 is that it can be used either on a confocal microscope or on a prism-based total internal reflection fluorescence microscope (TIRFM). Since excitation power, time resolution, and bleaching times are complementary for the two methods, this extends the usability of these surface immobilization experiments.

Cleaning of all the involved components is absolutely crucial for these experiments. Detailed below are some important considerations for the cleaning process.

Before the measurement:

- In general: Avoid touching of any of the components and wear nitrile gloves, cleaned with liquid soap prior to component handling.

- Before fitting into the holder, gaskets should be manually cleaned with clean-room swabs using a 3% Deconex solution, special care should be taken to clean the holes. After manually cleaning, sonicate gaskets 10 min in 3% Deconex and 10 min in 30% H_2O_2 at 60 °C. After multiple washing cycles with H_2O they are ready for mounting on the holder.

- Clean the teflon tweezers (shown in figure 37(A)) in hot 3% Deconex before manipulating the gaskets.

- Quartz glass, used to build the flow-cell chamber, should be cleaned according to the following protocol: sonicate for 10 min in 3% Deconex at 60 °C, sonicate for 10 min in 20% H_2O_2 solution (30%) + 20% HCl (32%) at 60 °C, sonicate for 10 min in 20% H_2O_2 solution (30%) + 20% ammonia solution (25%) at 60 °C, sonicate for 10 min in H_2O. Indicated in parenthesis are the stock concentrations of the respective solutions.

---

[20] Symphotime, PicoQuant GmbH, Berlin, Germany, (http://www.picoquant.com/)

[21] Free Serial Port Monitor, HHD Software, London, United Kingdom, (http://www.serial-port-monitor.com/)

[22] PCAN-View, PEAK-System Technik GmbH, Darmstadt, Germany, (http://www.peak-system.com/)

- If covalently functionalized cover glasses are used, appropriate cleaning protocols such as sonication in Tween 20 solutions should be applied.
- Use nitrogen gas from the tap to dry all components and to remove residual dust particles.

After the measurement

- For reuse, dissolve the quartz flow cell in acetone (at least over night).
- Mechanically clean the dissolved quartz slides using swabs with acetone, 3% Deconex, isopropanol and H_2O.
- Clean the gaskets mechanically using swabs with 3% Deconex.
- Purge all the tubings, syringes and connections with 3% Deconex and incubate everything in a bowl at 60 °C for 30 min (Avoid contact of Deconex with syringe pistons and the brass flow cell holder since base metals will be corroded by Deconex!). Next, purge everything again with hot 3% Deconex and then rigorously with H_2O. After drying with nitrogen gas, place everything on cleanroom towels and leave it to dry.
- Depending on the applied protein samples, there is the risk of residual aggregated products in the tubings or syringes. An additional cleaning step with 1M sodium hydroxide can be applied.
- Clean the brass flow-cell holder briefly with 3% Deconex and isopropanol. For the fittings, soak a swab and screw it trough the holes. Afterwards immediately rinse the holder with H_2O and dry it with nitrogen gas.

A general overview of the steps involved in assembling the flow cell and mounting it onto the confocal microscope is shown in figure 37. A more detailed view of the flow cell including the quartz sample cell can be found in figure 38.

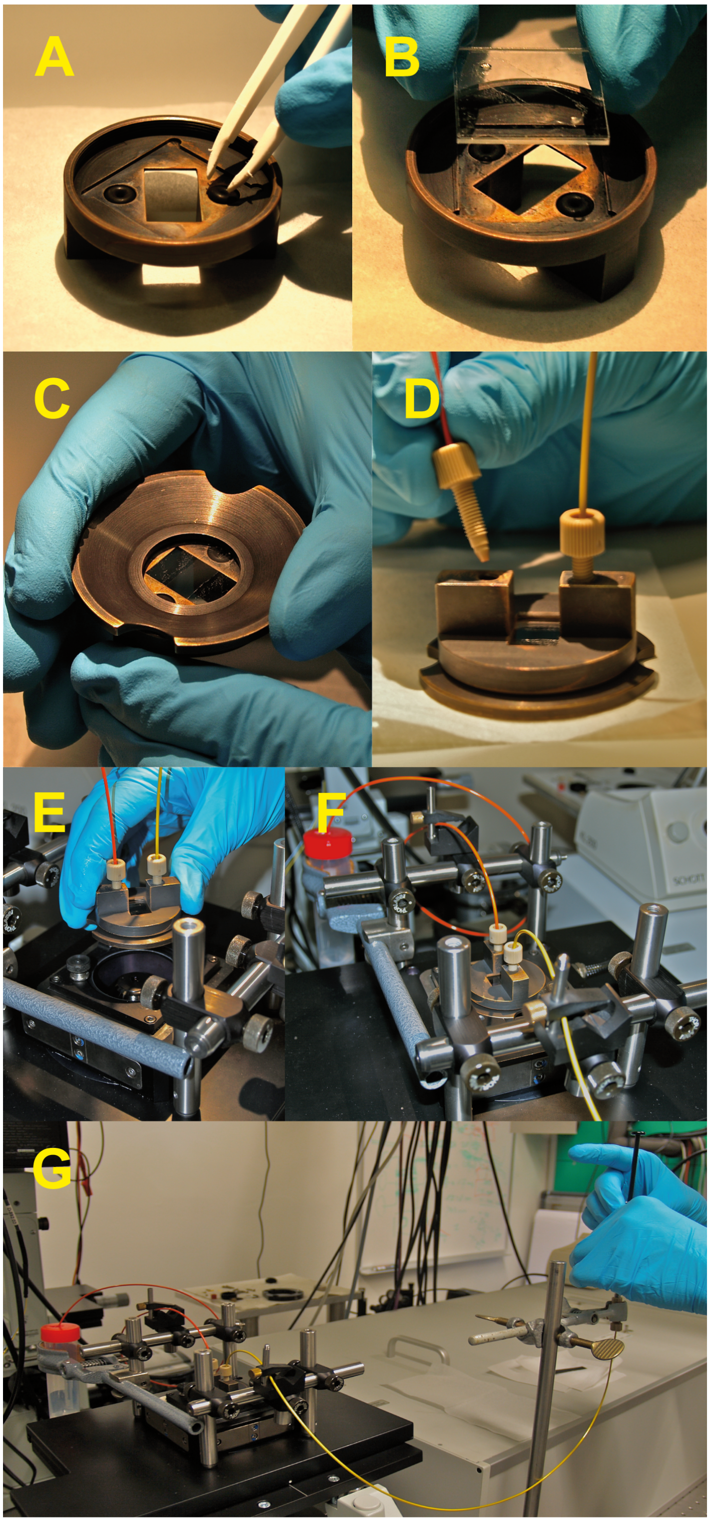


Figure 37: Assembling a flow cell for surface experiments. **A)** Fitting of the two gaskets into the metal holder. **B)** Insertion of the quartz-quartz flow-cell sandwich. **C)** Locking the flow-cell with the counter piece of the flow cell holder. **D)** Connecting the tubing to the flow-cell. **E)** Mounting the holder with the flow-cell on the confocal microscope. Due to the height of the flow-cell, an extension tube for the objective may be needed **F)** View of the mounted flow-cell with the secured capillaries to reduce vibrations, and the waste collection container in the back. **G)** Overview including the application syringe on the right.

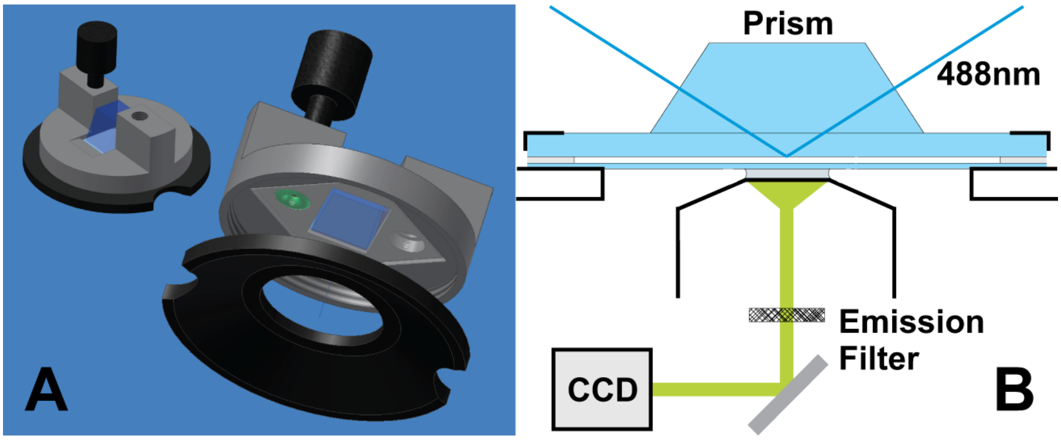


Figure 38: Schematic drawings of the flow-cell holder **A)** 3D representation of the flow-cell holder with an optional TIRFM prism in blue, a single gasket in green and a fitting in black. **B)** Sectional drawing of the flow-cell with prism and TIRFM excitation. The sample cell sandwich consisting of a thin and a thick cover glass (blue) connected by a layer of glue (gray). In the lower part, a schematic view of the TIRFM fluorescence collection part consisting of a high nummerical aperture objective, an emission filter and an EMCCD camera can be seen. Depending on the microscope setup, the cell can be used in TIRFM (with prism) or in confocal (without prism) mode. (Adapted from Haenni [32])

Surface immobilization of FRET labeled proteins on biotinylated PEG surfaces is a powerful and convenient method. Since the GdmCl stability and background properties are significantly better compared to immobilization on lipid bilayers, for most of the experiments this will be the method of choice. Described below is a short protocol of how such a surface experiment can be performed. If commercial, covalently bound biotin-PEG cover slides[23] are used, step two can be ignored.

1. Apply a 1:10 dilution (To reduce ionic strength but to still have some residual buffer capacity) of the buffer of choice into the flow cell and measure the background using Symphotime area scan. Usually 80x80 µm with a resolution of 250 nm is a good choice. The background should be measured at identical laser powers as the final measurement. Always close the laser shutter between measurements to avoid photobleaching.

2. Dissolve 0.1 mg/ml PLL(20)-g[3.5]-PEG(2)/PEG(3.4)-Biotin(50%)[24] in 150 µl of the diluted buffer, vortex, sonicate for 10 min and apply to the flow-cell. After 5-10 min incubation time, purge the flow-cell with 1 ml of undiluted buffer and make another background scan as described above.

3. Apply 1 mg/ml Avidin D[25] in buffer to the flow-cell. After 5-10 min incubation time, purge the flow-cell with 1 ml of undiluted buffer and make another background scan as described above. Load the area scan into Erntikon (See figure 39), adjust the contrast and use the "Automatic ROI Search" (See figure 39) with parameters as the ones planned to use for the single-molecule experiment. Reasonable values for the lower intensity threshold are between 10-20. Since all the located spots belong to the background, there should not be more than 30-60 ROI's be identified at this point.

4. If the fluorescence background is low enough, the flow-cell is now ready for sample application. Multiple sample applications are possible but there is no satisfactory strategy to

---

[23] Bio_01, MicroSurfaces Inc., Austin, TX, USA, (http://proteinslides.com/)

[24] Gift from Prof. Dr. Hans Marcus Textor, Swiss Federal Institute of Technology, Zurich, Switzerland, (Alternative Source: http://www.susos.com/)

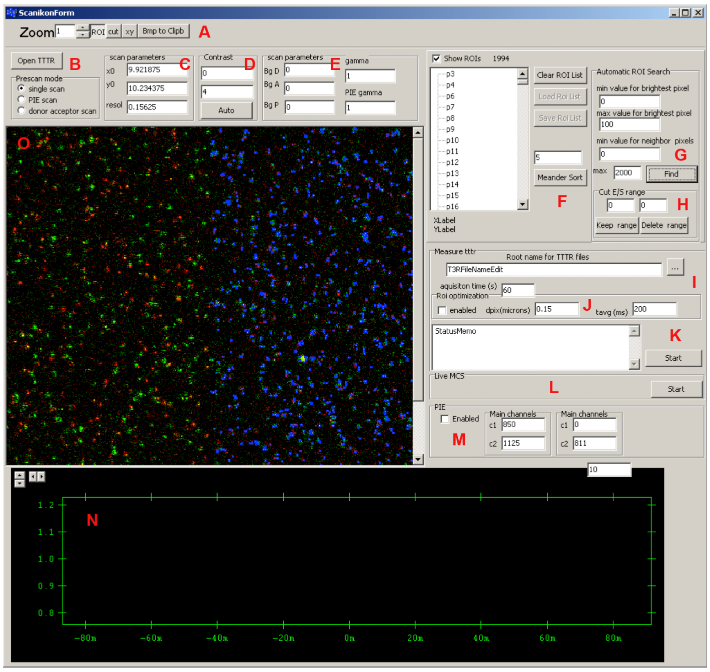
[25] A-2000, Vector Laboratories inc., Burlingame, CA, USA

remove immobilized molecules; cautious applications of low concentrations is advised. Usually concentrations of 10-50 pM of biotinylated sample protein are a good starting point. After every application it is advisable to select a new scan area to assess surface density. Since binding is usually quantitative, no purging between sample applications is needed. To avoid diffusive background events, purging with buffer after the last sample application is required. Depending on the photo-protection and fluorescence enhancement strategy, additives such as oxygen scavenging systems can now be applied.

5. Usually the stability should be high enough to allow measurements for hours to days. Buffer exchanges, such as to perform a GdmCl titration should be feasible without problems.

6. If high salt buffers, such as GdmCl solutions, were used the flow-cell should be disassembled and cleaned immediately after the measurements, since such solutions can catalyze the corrosion of the brass holder.

In rare cases, depending on the buffer conditions, swelling of the glue layer connecting the two quartz coverglasses may occur. To reduce this effect, the coverglasses can be activated in the plasma etcher immediately before gluing the two pieces together. Five minutes of etching with maximum power is recommended. This procedure leads to a very strong bonding of the flow cell which completely prevents dissociation and reuse of the coverglasses.



ScanikonForm

Zoom 1

Prescan mode: ● single scan ○ PIE scan ○ donor acceptor scan

scan parameters: x0 9.921875; y0 10.234375; resol 0.15625

Contrast: 0; 4

scan parameters: Bg D 0; Bg A 0; Bg P 0

gamma 1; PIE gamma 1

☑ Show ROIs 1994

5

Automatic ROI Search: min value for brightest pixel 0; max value for brightest pixel 100; min value for neighbor pixels 0; max 2000

Cut E/S range: 0; 0

Measure tttr: Root name for TTTR files T3RFileNameEdit; aquisiton time (s) 60

Roi optimization: ☐ enabled; dpix(microns) 0.15; tavg (ms) 200

StatusMemo

PIE: ☐ Enabled; Main channels c1 850, c2 1125; Main channels c1 0, c2 811

10

Figure 39: Screenshot of the software ERNTIKON developed by Daniel Nettels for the automated measurement of surface immobilized single molecules using a MT200 confocal microscope. ERNTIKON allows the loading and displaying (O) of "prescans" (measured before by using the area scan function of Symphotime). Immobilized particle positions (ROIs, regions of interest indicated by blue marks in (O)) can be identified manually or by using an automated search algorithm. Subsequently the identified particles can be measured one by one in an automated way. For this the piezo table is brought into position and a TTTR measurement is started. In the following some of the features and controls of ERNTIKON are described. **A)** Zoom level of the prescan and buttons for choosing the operation mode of the software: "ROI" enables the manual selection of region of interests (ROIs), "cut" for showing in panel (N) intensity profiles ("cuts") along horizontal lines chosen by clicking into the prescan , and "xy" for positioning the piezo table by clicking into the prescan. **B)** Opening of prescan TTTR files and choice of the prescan mode. **C)** Scan settings as obtained from the prescan TTTR file header. This information is needed for an exact positioning of the piezo table. In some rare cases fine adjustment of the values is needed. **D)** Contrast adjustment **E)** Depending on the prescan mode, one can set corresponding background and gamma values. This has an effect on (O) and the automatic ROI identification. **F)** Here the list of ROIs is shown and can be edited, saved or loaded. The ROI list defines the order of particle measurements. With "Meander Sort" this order can be optimized for avoiding large piezo table travels between measurements. **G)** ROI selection window. Based on threshold criteria ROIs are identified automatically. **H)** Transfer efficiency criterion applied to the ROI list. This can for example be used to avoid sampling of donor-only ROIs. Depending on the prescan mode (B) cut ranges either in transfer efficiency E or stochiometric ratio S are applied. **I)** Field to specify saving path, file name and acquisition time per "ROI". **J)** Local signal optimization. Software tries to maximize the counts by slightly varying the position near the ROI. **K)** "Start" button which initiates the sampling of the specified ROI's and status window that reports on the measuring progress. **L)** "Live MCS" starts a live-view of the detector signal displayed in (N). This can be used in combination with the "xy" mode to steer the confocal area manually to every position selected in (O) and to get a direct feedback on the signal collected there. The collected signal is not recorded to hard-disk. **M)** If the measurements are done in PIE mode, lifetime channel constraints can be entered here. **N)** Time trace of the fluorescence counts. The width of a time bin can be adjusted in the top right corner in units of milliseconds. Double-clicking the display opens a menu with various options for signal representation. **O)** Prescan display with some selected ROIs (blue boxes).

8.2 DNA as a Spectroscopic Ruler for Surface Experiments

The polyprolines generally used as a reference system for single-molecule experiments in solution are difficult to immobilize on surfaces. To have an additional system to test fluorescent dyes, substances for photo-enhancement and photo-protection and the effect of immobilization on the transfer efficiency histograms, a DNA system for use as a molecular ruler was developed. The design of the different variants can be found in figure 40. As can be seen in figure 40, there is a mother-strand with a biotin and non-paired, flexible linker. This strand can be annealed to different daughter-strands with different amino-dT positions and therefore different inter-dye separations, ranging from 10 to 30 base pairs. The sequence was chosen in such a way as to have a high stability, and therefore melting temperature, of the annealed double strand and a low self similarity which avoids secondary structure formation and shifted annealing. Before and after the labeling position an overhang was introduced to avoid unwinding of the helix close to the dye and to reduce possible stacking of the dyes to the ends of the helix. The 10 base pair distance separation between the different variants which approximately corresponds to a single turn was chosen in order to have the dyes on the same side of the helix. The 14 base pair variant was produced after the initial design and production process to have an additional variant with an intermediate transfer efficiency. That is also the reason why there is a base pair mismatch after annealing to the mother-strand. Due to the incorporation of the artificial nucleotide C6-dT it was possible to label the strands using N-hydroxy-succinimidyl ester functionalized dyes. The oligonucleotides were synthesized by solid-phase syhthesis[26] and delivered as desalted product. Ordering higher purity products turned out to be disadvantageous since the yields are much better by doing the purification for oneself. For the purification of the DNA, the following protocol was used:

1. Prepare the following buffers: **A)** 5M Urea, 12.5 mM Tris-HCl, pH 7.4 **B)** 5M Urea, 12.5 mM Tris-HCl, 0.5M Sodium Perchlorate, pH 7.4 (Filter the buffers carefully!).
2. Heat the DNAPac[27] ion exchange column at a flow rate of 0.1 ml/min to 85 °C.
3. Purify the unlabeled or labeled DNA with a gradient from 10-60% **B** in 30 min at 1 ml/min. Start the gradient with a time offset relative to the injection time for pre-gradient equilibration. Adjust the gradient after an analytical run according to your needs (Do not exceed 20 OD @ 260 nm per run).
4. Wash the column after each run with 100% **B** for 5 minutes.
5. The fractions of interest can be desalted using Sep-Pak cartridges[28] using the following steps:
6. Produce a 1M Triethylammonium Acetate (**TEAAc**) buffer at pH 7.5 (3g Acetic Acid and 5g Triethylamine in 50 ml H_2O).
7. Wash a new Sep-Pak column with 10 ml Acetonitrile followed by 10 ml of 0.1M **TEAAc** buffer.
8. Add 10% v/v of 1M **TEAAc** to the fraction of interest to get a final concentration of 0.1M **TEAAc** in the sample.
9. Apply the sample onto the prepared Sep-Pak cartridge.
10. Wash the cartidge with 10 ml 0.1M **TEAAc** followed by 10 ml H_2O. Dry the column using a syringe.

---

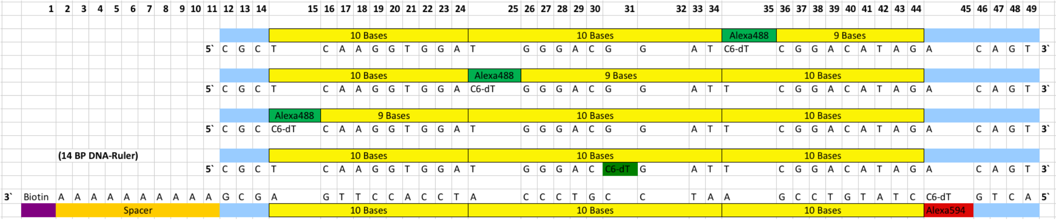
[26] Microsynth AG, Balgach, Switzerland

[27] DNAPac PA-100 (043010, 4x250 mm), Dionex (Switzerland) AG, Olten, Switzerland

[28] Sep-Pak Classic C18 cartridge, Waters Corporation, Milford, MA, USA

11. Elute the DNA with 3 ml 50% Acetonitrile in H_2O.
12. Lyophilize the sample.
13. After purification, remove the DNAPac column from the column heater, wash it for at least 10 min with 20% EtOH at 0.5 ml/min.

Several variants using different FRET dye pairs were produced and measured in solution and immobilized on surfaces. To verify the resolution in terms of transfer efficiencies in surface experiments, mixtures of variants with different inter-dye distances were applied on the same surface. It is important to mention that the ionic strength and composition of the buffer used for the measurements can have a significant influence on the expansion of the DNA helix and therefore on the measured transfer efficiency. Example histograms of two different DNA rulers can be found in figure 41.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49

10 Bases 10 Bases Alexa488 9 Bases
5` C G C T C A A G G T G G A T G G G A C G G A T C6-dT C G G A C A T A G A C A G T 3`

10 Bases Alexa488 9 Bases 10 Bases
5` C G C T C A A G G T G G A C6-dT G G G A C G G A T T C G G A C A T A G A C A G T 3`

Alexa488 9 Bases 10 Bases 10 Bases
5` C G C C6-dT C A A G G T G G A T G G G A C G G A T T C G G A C A T A G A C A G T 3`

(14 BP DNA-Ruler) 10 Bases 10 Bases 10 Bases
5` C G C T C A A G G T G G A T G G G A C C6-dT G A T T C G G A C A T A G A C A G T 3`

3` Biotin A A A A A A A A A A G C G A G T T C C A C C T A C C C T G C C T A A G C C T G T A T C C6-dT G T C A 5`
Spacer 10 Bases 10 Bases 10 Bases Alexa594

Figure 40: Design and sequence of the spectroscopic DNA Ruler. The amino group of the C6-dT in each sequence can be labeled using N-hydroxy-succinimide chemistry. The first four shorter sequences represent the daughter-strands which can be choosen depending on the distance needs. The strand labeled “14 BP DNA-Ruler” was produced after the initial design process to have an intermediate distance between 10 and 20 BP. That is also the reason why it carries a base pair mismatch after annealing to the mother-strand. The last sequence represents the mother-strand with the 3‘ attached biotin for surface immobilization.

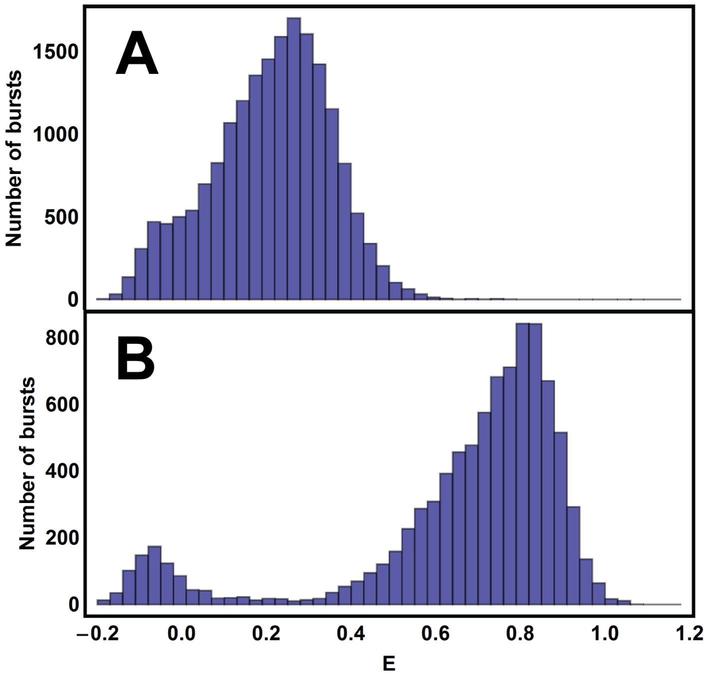


Figure 41: Example transfer efficiency histograms of DNA rulers labeled with Alexa 488 as donor and Alexa 594 as acceptor fluorescent dyes. **A)** Histogram of the 20 base pair ruler. **B)** Histogram of the 10 base pair ruler.

8.3 Data Analysis for TIRFM Experiments

Since photobleaching of the fluorophores in immobilized surface experiments usually takes longer with TIRFM compared to confocal measurements, these methods can be used complementarily. Additionally, there is the advantage that one can observe more than one molecule at a time in TIRFM which can be a significant asset when performing kinetic experiments. In contrast to confocal experiments, the data collection in TIRFM experiments is performed by an EMCCD camera. The collected fluorescence image is split by a dichroic mirror according to the colors of the donor and the acceptor photons and directed onto two different areas of the CCD chip. By overlaying the two images, it is possible to reconstruct the transfer efficiency of the molecules on the surface. By processing a movie, instead of a single image, it is possible to aquire temporal information. The alignment of the donor and the acceptor images can be complicated due to slight variations in size and orientation. Therefore a novel strategy using MATLAB[29] and the Image Processing Toolbox[30] was developed. This software environment was chosen because of the built-in image manipulation functions and its rather high performance on such data structures. In the following section the most important functions and their application is described.

Due to limitations in memory allocation, it is required to convert the .fits files generated by the EMCCD camera software to .tiff image stacks. These can then be imported frame by frame, which reduces the required memory. To best be able to present the code here, some additional line breaks were introduced.

---

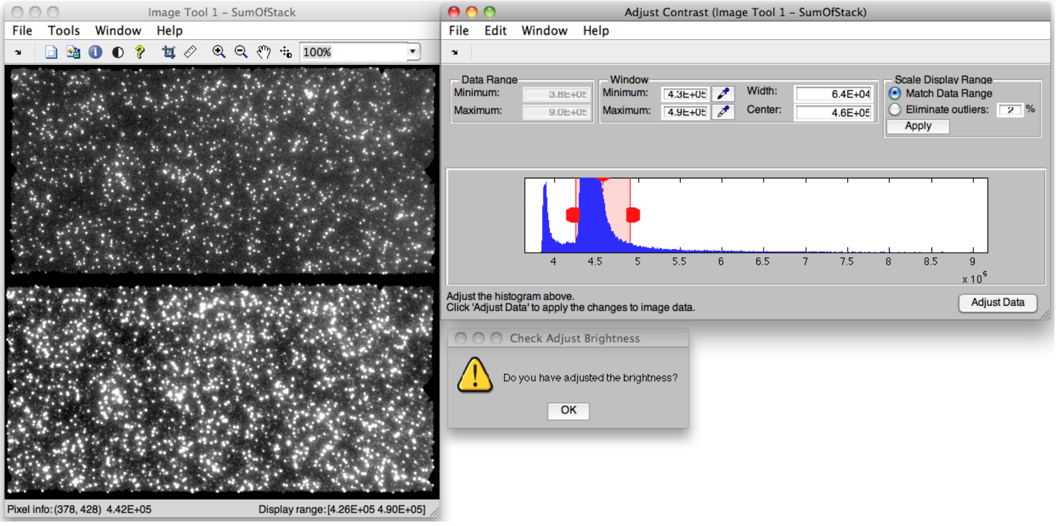
[29] R2008b, The MathWorks Inc., Natick, MA, USA

[30] V6.2, The MathWorks Inc., Natick, MA, USA

Opening a .tiff stack

```
function [ DStack,AStack,DSum,ASum ] = tiff2stack(  )
%FITS2STACK Summary of this function goes here
%Detailed explanation goes here
[filename, pathname]=uigetfile('*.tif', 'Select your .tif Stack');
tiffname=fullfile(pathname,filename);

NumberOfImages=length(imfinfo(tiffname));

for k = 1:NumberOfImages
I{k} = imread(tiffname, k);
end

SumOfStack=double(I{1});

for k=2:NumberOfImages
SumOfStack=SumOfStack+double(I{k});
end

AdjustBrightness=imtool(SumOfStack);
imcontrast(AdjustBrightness);
uiwait(warndlg('Do you have adjusted the brightness?','Check Adjust Brightness'))
BrightnessAdjusted=getimage(AdjustBrightness);

%Achtung: TIRFIKON dreht Kamera-Bilder um!

ASum=imcrop(BrightnessAdjusted,[1 257 512 555]);
DSum=imcrop(BrightnessAdjusted,[1 1 512 255]);

clear SumOfStack;

h = waitbar(0,'Calculating Acceptor-Stack: Please wait...');
for k=1:NumberOfImages
    waitbar(k/NumberOfImages)
    AStack(:,:,k)=single(I{k}(257:512,1:512));
end
close(h)

h = waitbar(0,'Calculating Donor-Stack: Please wait...');
for k=1:NumberOfImages
    waitbar(k/NumberOfImages)
    DStack(:,:,k)=single(I{k}(1:256,1:512));
end
close(h)
clear I;

end
```

Figure 42: tiff2stack matlab code: Opens a .tif stack, adjusts the brightness of the summed image stacks to allow identification of donor and acceptor spots, cuts out and produces a 3D data-structure for the donor and acceptor signal.



Data Range
Minimum: 3.8E+05
Maximum: 9.0E+05

Window
Minimum: 4.3E+05
Maximum: 4.9E+05
Width: 6.4E+04
Center: 4.6E+05

Scale Display Range
● Match Data Range
○ Eliminate outliers: 2 %

Adjust the histogram above.
Click 'Adjust Data' to apply the changes to image data.

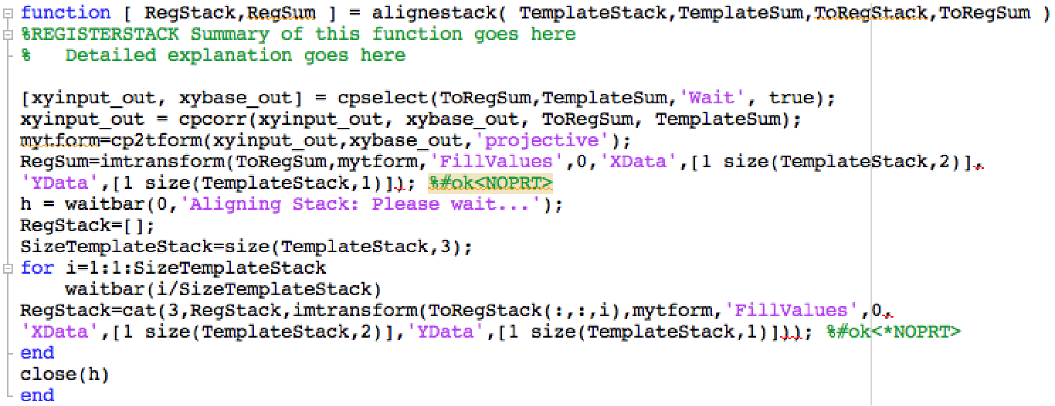
Do you have adjusted the brightness?

Pixel info: (378, 428) 4.42E+05

Display range: [4.26E+05 4.90E+05]

Figure 43: tiff2stack contrast adjustment example

Alignment of the donor and acceptor stacks



```
function [ RegStack,RegSum ] = alignestack( TemplateStack,TemplateSum,ToRegStack,ToRegSum )
%REGISTERSTACK Summary of this function goes here
%   Detailed explanation goes here

[xyinput_out, xybase_out] = cpselect(ToRegSum,TemplateSum,'Wait', true);
xyinput_out = cpcorr(xyinput_out, xybase_out, ToRegSum, TemplateSum);
mytform=cp2tform(xyinput_out,xybase_out,'projective');
RegSum=imtransform(ToRegSum,mytform,'FillValues',0,'XData',[1 size(TemplateStack,2)],
'YData',[1 size(TemplateStack,1)]); %#ok<NOPRT>
h = waitbar(0,'Aligning Stack: Please wait...');
RegStack=[];
SizeTemplateStack=size(TemplateStack,3);
for i=1:1:SizeTemplateStack
    waitbar(i/SizeTemplateStack)
RegStack=cat(3,RegStack,imtransform(ToRegStack(:,:,i),mytform,'FillValues',0,
'XData',[1 size(TemplateStack,2)],'YData',[1 size(TemplateStack,1)])); %#ok<*NOPRT>
end
close(h)
end
```

Figure 44: Alignestack matlab code: opens an interface window where identical points in the donor and the acceptor window can be manually identified. After local optimization of these anchoring points using crosscorrelation, the coordinate systems of the two image stacks are overlaid.

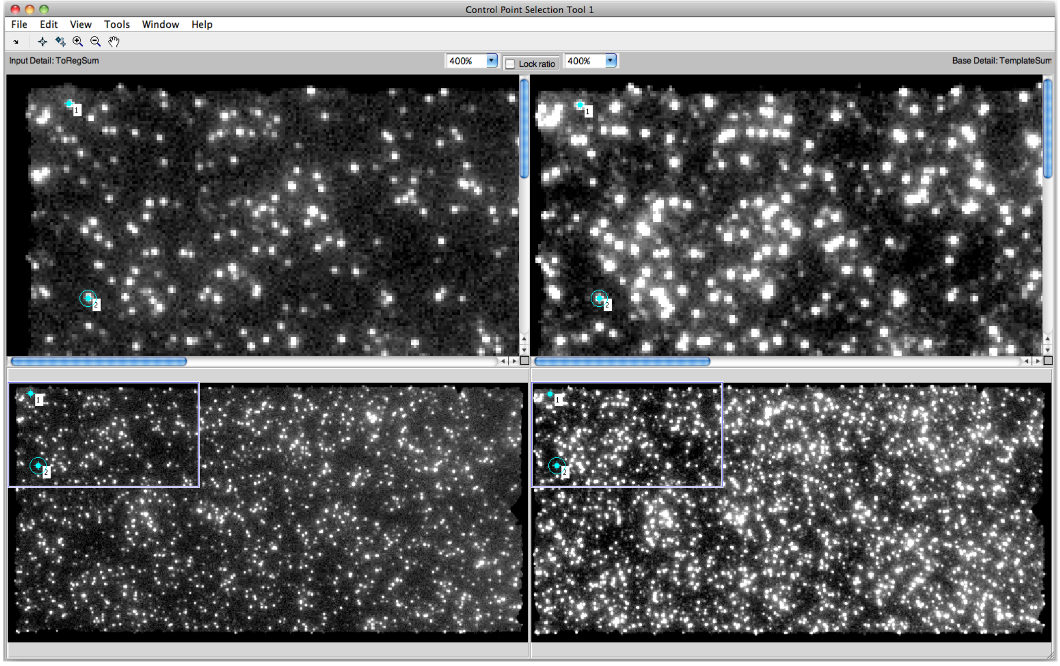
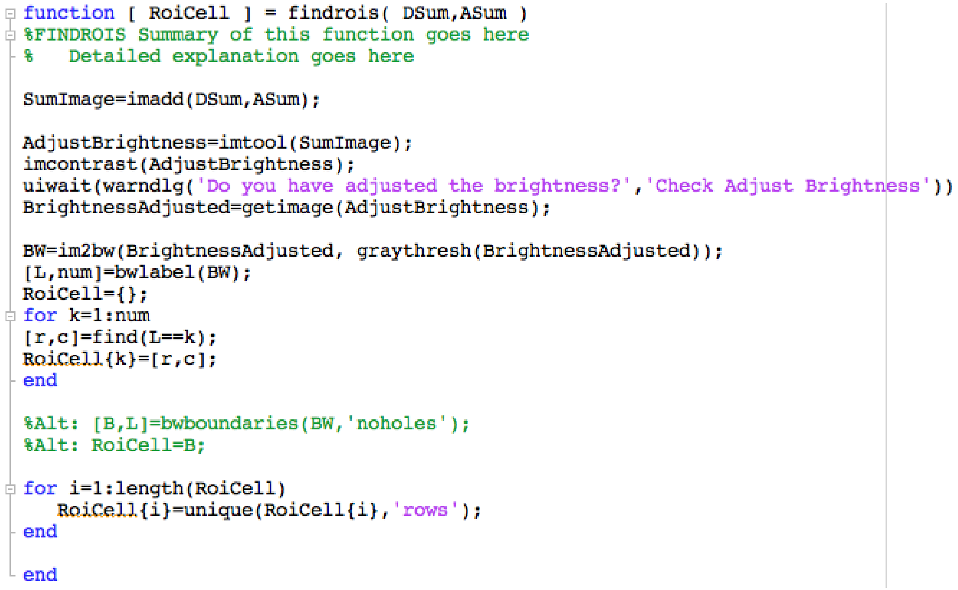


Figure 45: Identification of identical points in the donor and the acceptor image using the alignestack function.

Find the ROI‘s



```
function [ RoiCell ] = findrois( DSum,ASum )
%FINDROIS Summary of this function goes here
%   Detailed explanation goes here

SumImage=imadd(DSum,ASum);

AdjustBrightness=imtool(SumImage);
imcontrast(AdjustBrightness);
uiwait(warndlg('Do you have adjusted the brightness?','Check Adjust Brightness'))
BrightnessAdjusted=getimage(AdjustBrightness);

BW=im2bw(BrightnessAdjusted, graythresh(BrightnessAdjusted));
[L,num]=bwlabel(BW);
RoiCell={};
for k=1:num
[r,c]=find(L==k);
RoiCell{k}=[r,c];
end

%Alt: [B,L]=bwboundaries(BW,'noholes');
%Alt: RoiCell=B;

for i=1:length(RoiCell)
    RoiCell{i}=unique(RoiCell{i},'rows');
end

end
```

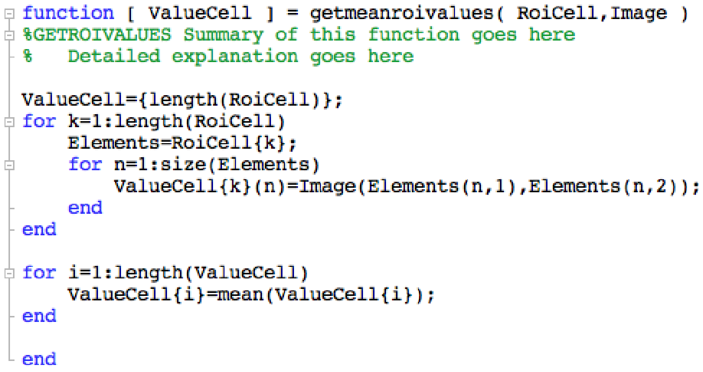
Figure 46: Findrois matlab code: Using the same contrast adjustment window shown in figure 43 the user has to adjust the thresholds of the overlaid donor/acceptor images in such a way that the background is black and the ROI‘s are white. Based on this adjustment, a black and white image is generated and neighboring white pixels are united in single ROI‘s. A data-structure with the coordinates of all the pixels in each ROI is returned.

Find the local background

```
function [ BackgroundPixelCell ] = newfindbackgroundrois(RoiCell,PixelOffset,RegSumImage)
%GETBACKGROUND Summary of this function goes here
%   Detailed explanation goes here

BackgroundPixelCell={};
LengthMax=size(RegSumImage,2);
WidthMax=size(RegSumImage,1);
RoiMatrix=[];
p=0;
for i=1:length(RoiCell)
    for k=1:size(RoiCell{i},1)
        p=p+1;
        RoiMatrix(p,:)=RoiCell{i}(k,:);
    end
end

for k=1:length(RoiCell)
    maxVal=max(RoiCell{k},[],1);
    minVal=min(RoiCell{k},[],1);
    UpLe=minVal-PixelOffset;
    UpRi=[maxVal(1)+PixelOffset minVal(2)-PixelOffset];
    LoLe=[minVal(1)-PixelOffset maxVal(2)+PixelOffset];
    LoRi=maxVal+PixelOffset;

    p=0;
    BackgroundAreaArray=[];
    for i=UpLe(2):LoLe(2) %Iteration über y
        for g=UpLe(1):UpRi(1)
            p=p+1;
            BackgroundAreaArray(p,:)=[g i];
        end
    end
    p=0;
    WorkingMatrixNew=[];
    for i=1:size(BackgroundAreaArray,1)
        if BackgroundAreaArray(i,1)>0 && BackgroundAreaArray(i,2)>0 &&
            BackgroundAreaArray(i,1)<WidthMax+1 && BackgroundAreaArray(i,2)<LengthMax+1,
            p=p+1;
            WorkingMatrixNew(p,:)=BackgroundAreaArray(i,:);
        end
    end

    WorkingMatrixNewer=setdiff(WorkingMatrixNew,RoiMatrix,'rows');
    BackgroundPixelCell{k}=WorkingMatrixNewer;

end
end
```

Figure 47: Findbackgroundrois matlab code: Due to the possibility of an inhomogeneous distribution of background fluorescence, a local background taken around the ROI was introduced. This function generates a coordinate data structure where the position of the background pixels around each ROI is registered.

Calculate the mean of the ROI‘s and the background values



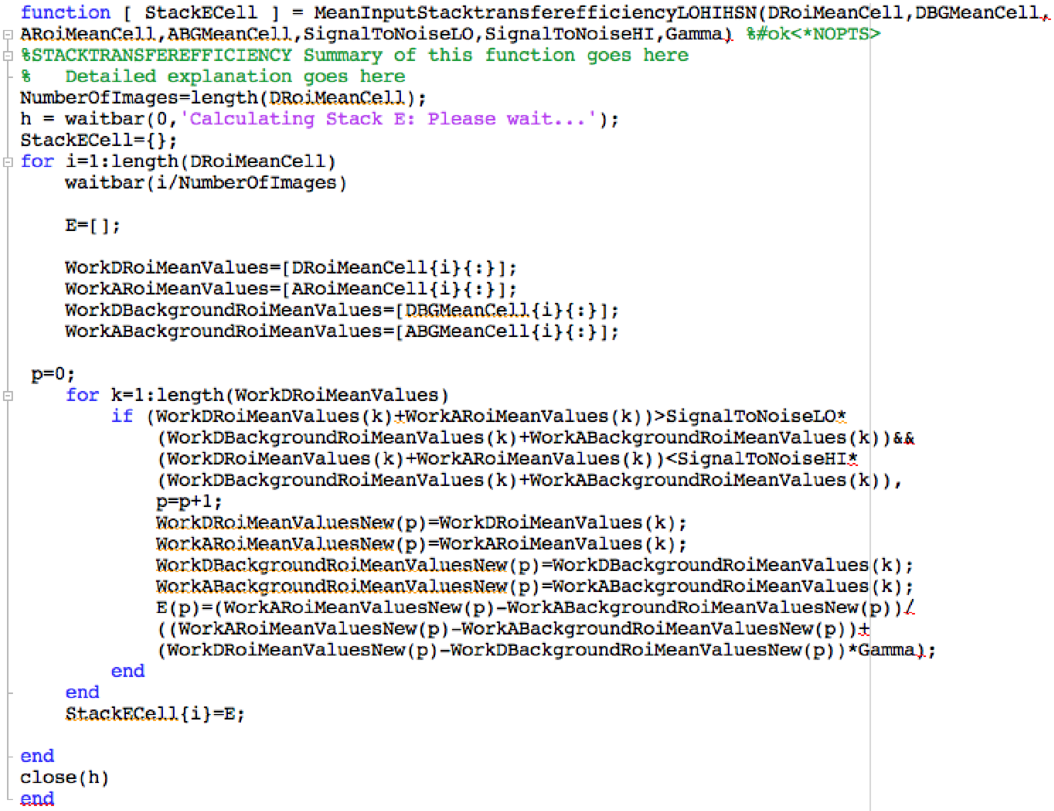
```
function [ ValueCell ] = getmeanroivalues( RoiCell,Image )
%GETROIVALUES Summary of this function goes here
%   Detailed explanation goes here

ValueCell={length(RoiCell)};
for k=1:length(RoiCell)
    Elements=RoiCell{k};
    for n=1:size(Elements)
        ValueCell{k}(n)=Image(Elements(n,1),Elements(n,2));
    end
end

for i=1:length(ValueCell)
    ValueCell{i}=mean(ValueCell{i});
end

end
```

Figure 48: For speed improvement, accessing the 3D data structures and calculating the mean of the pixel values within a ROI is packed in the function called getmeanroivalues.

```
function [ DRoiMeanCell,ARoiMeanCell,DBGMeanCell,ABGMeanCell ] =
MeanRoiGenerator( DStack,AStack,ROICell,BGROICELL)
%STACKTRANSFEREFFICIENCY Summary of this function goes here
%   Detailed explanation goes here
NumberOfImages=size(DStack,3);
h = waitbar(0,'Calculating Mean-Stack: Please wait...');

DBGMeanCell={NumberOfImages};
DRoiMeanCell={NumberOfImages};
ABGMeanCell={NumberOfImages};
ARoiMeanCell={NumberOfImages};

for i=1:NumberOfImages
    waitbar(i/NumberOfImages)

    WorkDStackValues=DStack(:,:,i);
    WorkAStackValues=AStack(:,:,i);

    DRoiMeanCell{i}=getmeanroivalues(ROICell,WorkDStackValues);
    ARoiMeanCell{i}=getmeanroivalues(ROICell,WorkAStackValues);
    DBGMeanCell{i}=getmeanroivalues(BGROICELL,WorkDStackValues);
    ABGMeanCell{i}=getmeanroivalues(BGROICELL,WorkAStackValues);

end
close(h)
end
```

Figure 49: Using the ROI and background coordinates data structures, the donor and acceptor mean signal and mean background are calculated for each ROI using the MeanRoiGenerator function.

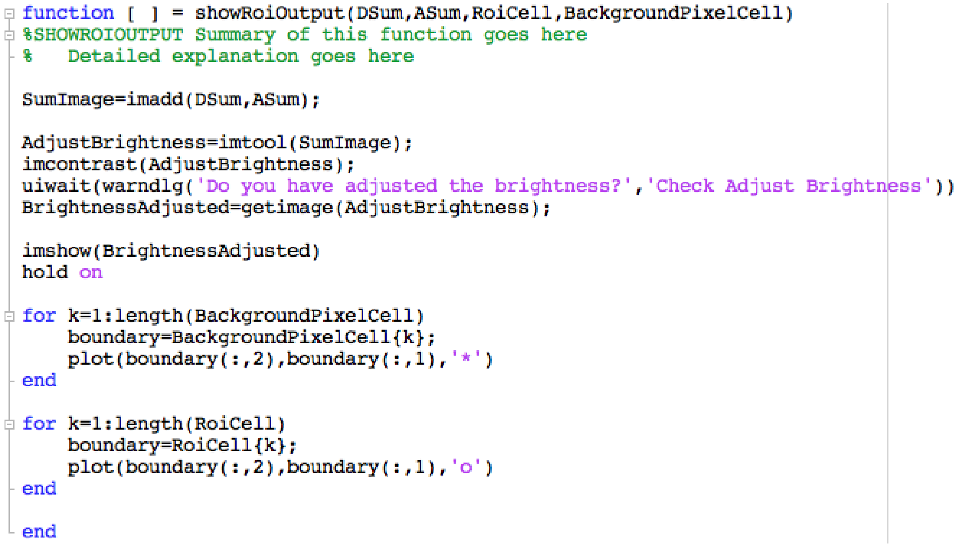
Calculate the transfer efficiencies



```
function [ StackECell ] = MeanInputStacktransferefficiencyLOHIHSN(DRoiMeanCell,DBGMeanCell,
ARoiMeanCell,ABGMeanCell,SignalToNoiseLO,SignalToNoiseHI,Gamma) %#ok<*NOPTS>
%STACKTRANSFEREFFICIENCY Summary of this function goes here
%   Detailed explanation goes here
NumberOfImages=length(DRoiMeanCell);
h = waitbar(0,'Calculating Stack E: Please wait...');
StackECell={};
for i=1:length(DRoiMeanCell)
    waitbar(i/NumberOfImages)

    E=[];

    WorkDRoiMeanValues=[DRoiMeanCell{i}{:}];
    WorkARoiMeanValues=[ARoiMeanCell{i}{:}];
    WorkDBackgroundRoiMeanValues=[DBGMeanCell{i}{:}];
    WorkABackgroundRoiMeanValues=[ABGMeanCell{i}{:}];

 p=0;
    for k=1:length(WorkDRoiMeanValues)
        if (WorkDRoiMeanValues(k)+WorkARoiMeanValues(k))>SignalToNoiseLO*
            (WorkDBackgroundRoiMeanValues(k)+WorkABackgroundRoiMeanValues(k))&&
            (WorkDRoiMeanValues(k)+WorkARoiMeanValues(k))<SignalToNoiseHI*
            (WorkDBackgroundRoiMeanValues(k)+WorkABackgroundRoiMeanValues(k)),
            p=p+1;
            WorkDRoiMeanValuesNew(p)=WorkDRoiMeanValues(k);
            WorkARoiMeanValuesNew(p)=WorkARoiMeanValues(k);
            WorkDBackgroundRoiMeanValuesNew(p)=WorkDBackgroundRoiMeanValues(k);
            WorkABackgroundRoiMeanValuesNew(p)=WorkABackgroundRoiMeanValues(k);
            E(p)=(WorkARoiMeanValuesNew(p)-WorkABackgroundRoiMeanValuesNew(p))/
            ((WorkARoiMeanValuesNew(p)-WorkABackgroundRoiMeanValuesNew(p))+
            (WorkDRoiMeanValuesNew(p)-WorkDBackgroundRoiMeanValuesNew(p))*Gamma);
        end
    end
    StackECell{i}=E;

end
close(h)
end
```

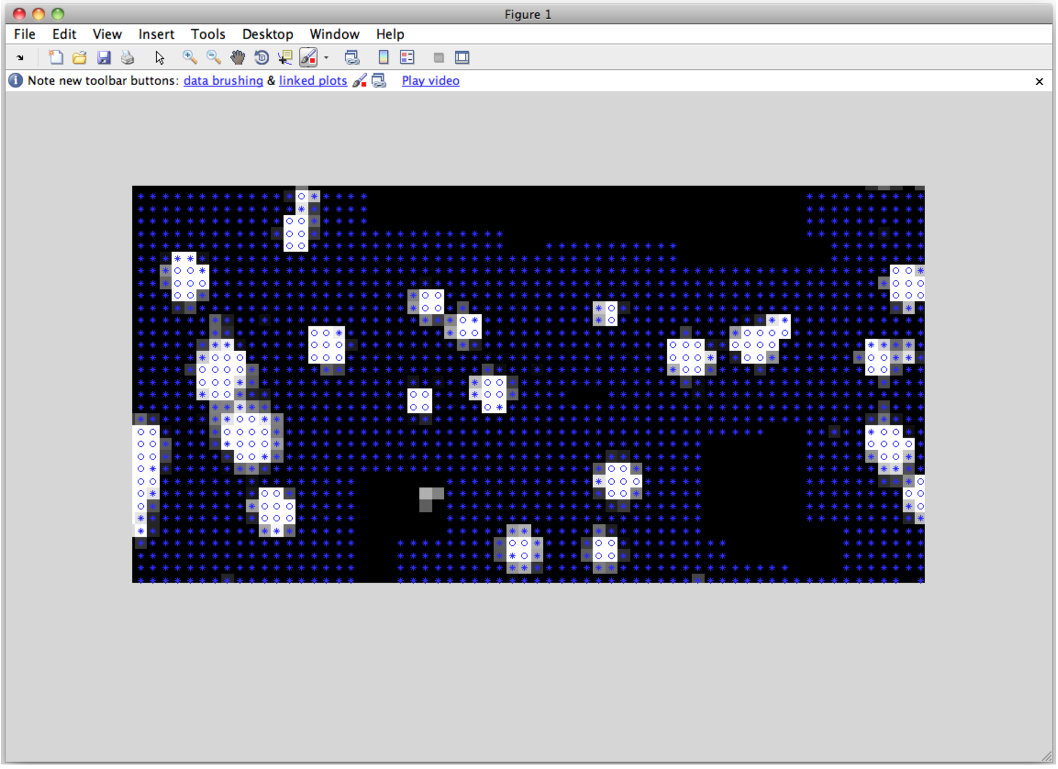
Figure 50: Using a signal threshold criterion and a gamma constant to correct for differences in donor and acceptor detection efficiencies, this function calculates the background corrected transfer efficiency for each ROI.

Show the identified ROI‘s and background areas



```
function [ ] = showRoiOutput(DSum,ASum,RoiCell,BackgroundPixelCell)
%SHOWROIOUTPUT Summary of this function goes here
%   Detailed explanation goes here

SumImage=imadd(DSum,ASum);

AdjustBrightness=imtool(SumImage);
imcontrast(AdjustBrightness);
uiwait(warndlg('Do you have adjusted the brightness?','Check Adjust Brightness'))
BrightnessAdjusted=getimage(AdjustBrightness);

imshow(BrightnessAdjusted)
hold on

for k=1:length(BackgroundPixelCell)
    boundary=BackgroundPixelCell{k};
    plot(boundary(:,2),boundary(:,1),'*')
end

for k=1:length(RoiCell)
    boundary=RoiCell{k};
    plot(boundary(:,2),boundary(:,1),'o')
end

end
```

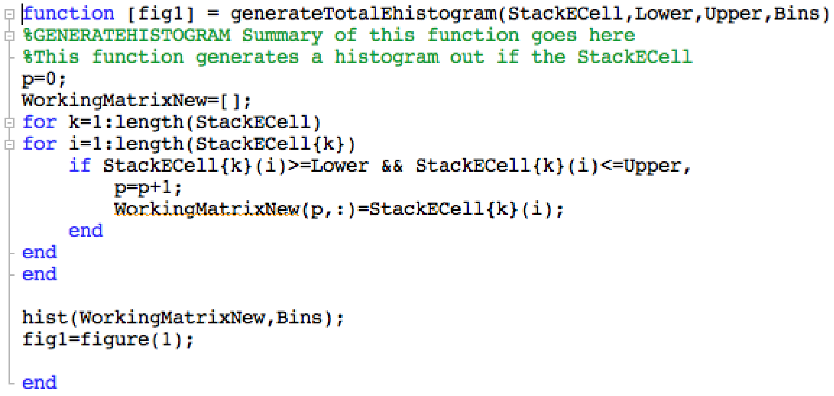
Figure 51: The showRoiOutput function, generates a graphical representation of the identified ROI‘s and the corresponding background pixels. This can be used to check the functionality of the code and to check if the if the thresholding is reasonable.



Note new toolbar buttons: data brushing & linked plots Play video

Figure 52: An example of the graphical output delivered by the showRoiOutput function. Circles represent ROI pixels and stars the corresponding background area. A problem of the current implementation can be seen in this picture. Since there are no security marigins between the ROI‘s and the corresponding background area, there is a certain chance that low intensity ROI pixels are attributed background area. This should be modified in a future application.

Generate the transfer efficiency histogram



```
function [fig1] = generateTotalEhistogram(StackECell,Lower,Upper,Bins)
%GENERATEHISTOGRAM Summary of this function goes here
%This function generates a histogram out if the StackECell
p=0;
WorkingMatrixNew=[];
for k=1:length(StackECell)
for i=1:length(StackECell{k})
    if StackECell{k}(i)>=Lower && StackECell{k}(i)<=Upper,
        p=p+1;
        WorkingMatrixNew(p,:)=StackECell{k}(i);
    end
end
end

hist(WorkingMatrixNew,Bins);
fig1=figure(1);

end
```

Figure 53: Matlab code to generate the transfer efficiency histogram.

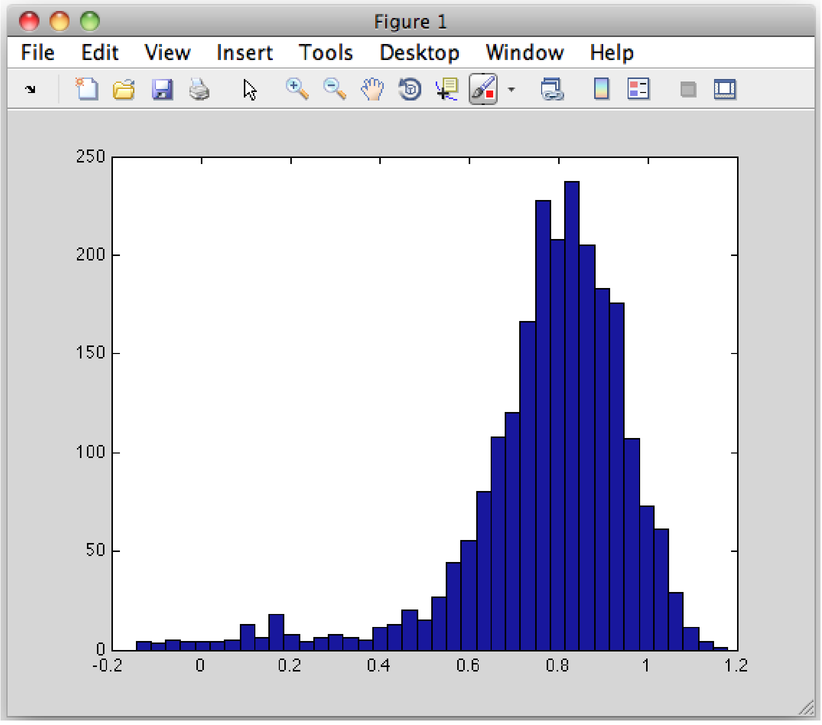


Figure 54: Output produced by the generateTotalEhistogram function. As an example, data from a DNA ruler immobilized on a biotinylated PEG surface was used.

8.4 Photo-Protection and Fluorescence-Enhancement Strategies

8.4.1 Substance Candidates

Based on an extensive search in chemical databases and literature, candidates for photoprotection and fluorescence enhancement were selected for single-molecule test experiments. Selection rationals were compunds with known effects on fluorescence bleaching or triplet state dynamics, radical scavengers, redox active compounds, preservatives or soluble substances with exposed electron systems which can act as triplet state quenchers. Colorful or fluorescent substances were removed from the selection in most cases since they are most probably unsuitable for single-molecule studies. For this reason, most of the polyphenol candidates had to be removed from the selection. To get more information about mechanisms and substance candidates, the reader is referred to the corresponding literature [125, 100, 19, 89, 124, 1, 119, 11, 13, 84]. A list of tested substance candidates can be found in table 4.

Substance	**CAS**	**Supplier**	**Suitable**
2,4,5-Trihydroxybutrophenone	1421-63-2	SU:47192	Yes
2-Mercaptoethanol	60-24-2	FL:63689	Yes
2,6-di-tert-butyl-4-(hydroxymethyl)phenol (Ionox 100)	88-26-6	SU:47192	Yes
4-Nitrobenzyl alcohol	619-73-8	AL:N1,282-1	Yes
Acetylcysteine	616-91-1	SI:A8199	Yes
Chloramphenicol	56-75-7	FL:23275	No
Cinnamic acid	140-10-3	FL:96340	Yes
Cysteamine	156-57-0	SI:M6500	Yes
Triethylenediamine (DABCO)	280-57-9	AL:D2,780-2	Yes
Diethylhydroxylamine (DEHA)	3710-84-7	AL:471593	Yes
Epigallocatechin gallate	989-51-5	FL:50299	No
Ethylenediaminetetraacetic acid (EDTA)	60-00-4	BS:05142391	Yes
Gallic acid	5995-86-8	SI:27645	No
Glutathione oxidized	27025-41-8	FL:49740	Yes
Glutathione reduced	70-18-8	SI:G4251	Yes
Hydroquinone	123-31-9	RH:15616	Yes
Imidazole	288-32-4	RO:3892.2	No
Lactic acid	79-33-4	SI:L6402	Yes
Nitrous oxide	10024-97-2	PG:610	Yes
Nordihydroguaiaretic acid (NDGA)	500-38-9	SU:47192	No
Methylviologen (Paraquat)	1910-42-5	FL:36541	Yes
Phloroglucinol	108-73-6	FL:79330	No
Butylpyridyloxidenitrone (POBN)	66893-81-0	SI:P9271	No
Propyl gallate	121-79-9	SU:47192	Yes
L-Selenomethionine	3211-76-5	SI:S3132	Yes
Sodium ascorbate	134-03-2	FL:11140	Yes
Sodium azide	26628-22-8	SI:438456	Yes
Sodium sulfite	7757-83-7	SI:71988	Yes
tert-Butylhydroquinone (TBHQ)	1948-33-0	SU:47192	Yes
Trisodium citrate	6132-04-3	SI:S4641	Yes
Trolox	53188-07-1	AL:23,881-3	Yes

Table 4: The substances used for the photo-protection and fluorescence enhancement screen. In the first column the chemical name of the substance can be found. In the second column, for rapid identification, the CAS registry number can be found. The third colum indicates the supplier and the product number for easier identification of the substance tested in this assay. The suppliers are coded by the following abbreviations: **SU**: Supelco, **FL**: Fluka, **SI**: Sigma, **AL**: Aldrich, **BS**: Biosolve, **RH**: Riedel-de Häen, **RO**: Roth, **PG**: PanGas. The last column states if the color, purity and fluorescence characteristics of the substance allowed a single-molecule measurement for the conditions used in this screen. However, it does not state if there was an improvement upon adding the specific substance.

In addition to the proposed additives the effect of the surfactant used in single-molecule experiments which is required to avoid adsorption of the proteins to the quartz walls of the measuring cell was also measured. The following surfactants were tested: NP-40, TX-100, Brij-35, Brij-58, TX-114, Tween-20. No difference in fluorescence was found by changing between these surfactants.

To evaluate the substances, the following strategy was used. First the substance was dissolved or diluted into a standard PBS buffer and checked for background fluorescence at various concentrations. Only substances with a low background were included in the following screen. Next a test peptide with Alexa 488 and Alexa 594 fluorescent dyes such as a labeled polyproline was used to test the influence of the substance at a single-molecule level. Various concentrations and combinations of substances were screened. To get more insights about possible photo-protective capabilities, the screen was generally performed at multiple laser powers. From the number of bursts obtained and the shape of the transfer efficiency histograms it was already possible to get a first impression about possible effects. Further analysis such as FCS lead to insights about triplet state population. From burst-size and burst-time distributions and photon counting histograms it was possible to get information about the molecular fluorescence brightness of single molecules. Together with the distribution of green and red photons within a burst, this data also reports on the occurrence of bleaching events. These experimental results allowed us to get a conclusive rating of the photo-protection and fluorescence-enhancement capabilities of the tested substance or of combinations of substances. Since single molecules are observed for a significantly longer time in surface experiments compared to diffusion experiments, photo-protection results in a significant expansion in the experimentally accessible timescale. Therefore, additional surface experiments using the DNA ruler presented in section 8.2 were performed for selected substance candidates. Since oxidation of the fluorophores seems to be a common pathway of photobleaching, various experiments using enzymatic oxygen depletion systems were performed. This usually led to a almost complete loss of the fluorescence signal for the Alexa 488 Alexa 594 dye pair, most probably due to the introduced lack of triplet state quenching by oxygen. Adding triplet state quenchers such as mercaptoethanol or redox cascades such as a mixture of ascorbic acid together with methylviologen were not able to fully recover the signal.

8.4.2 Automated Single-Molecule Experiment

To perform a large scale screen of concentrations, laser powers and substance combinations an automated sampling system, using single-use wellplates[31] with a glass bottom, was developed. Figure 55 shows the drawings of the holder produced to mount the wellplates on the confocal microscope. Large plate to plate differences in terms of planarity required the introduction of adjustment screws to align each wellplate on the microscope stage.

---

[31] Glass Bottomed Microplates, PS96B-G175, SWISSCI, Neuheim, Switzerland (http://www.swissci.com/)

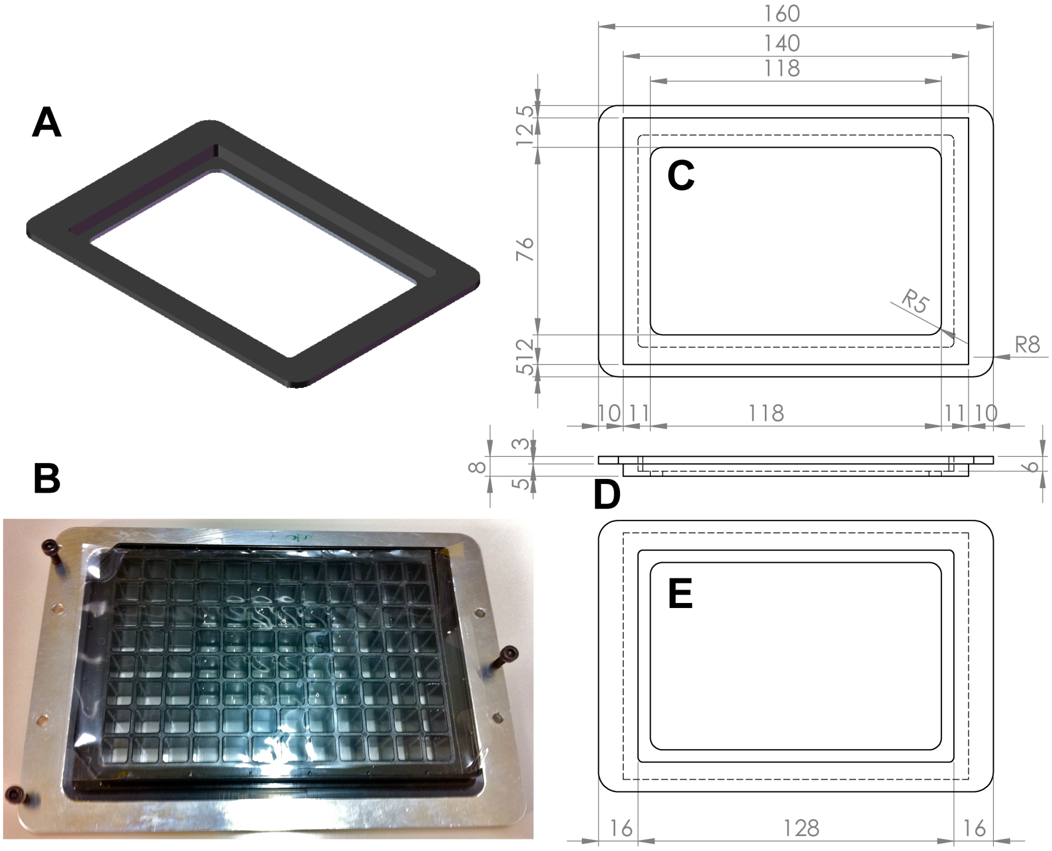


Figure 55: Image and drawings of the wellplate holder. **A)** Rendered 3D view of the holder. **B)** Image of the holder together with a sealed wellplate. The holes and screws for height adjustment can be seen. **C)** Bottom view. **D)** Side view. **E)** Top view. (Drawings by Bengt Wunderlich)

The automated stage with a mounted wellplate is shown in figure 56. The 36 outermost wells should not be used, or if absolutely necessary, some of the remaining wells may be used with great care, by positioning the objective at the innermost corner of these additional wells.

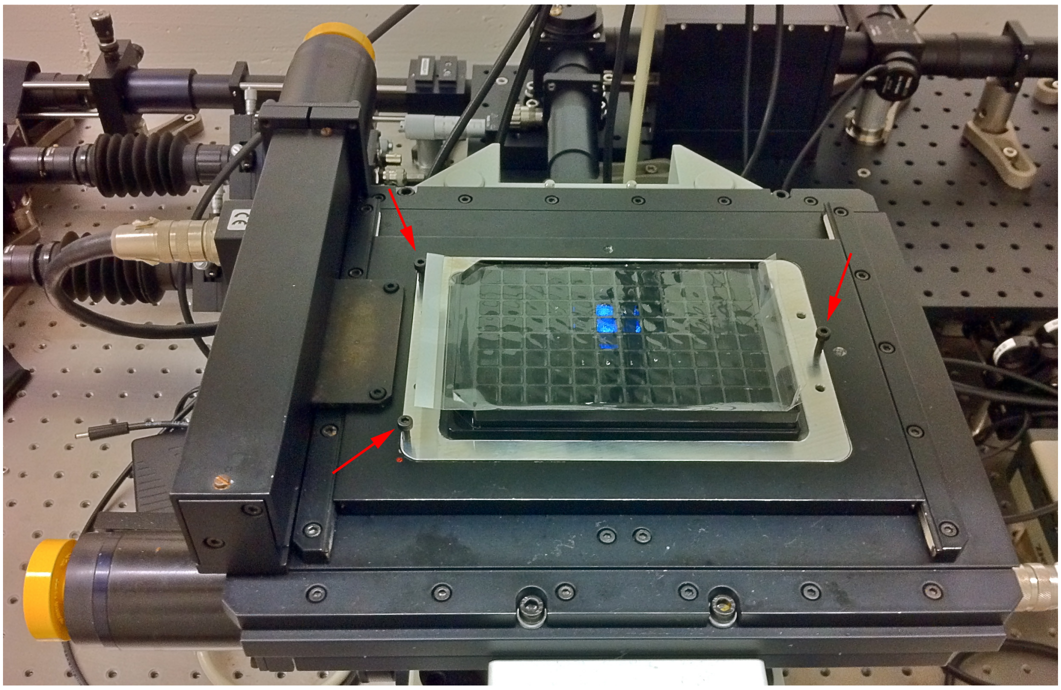
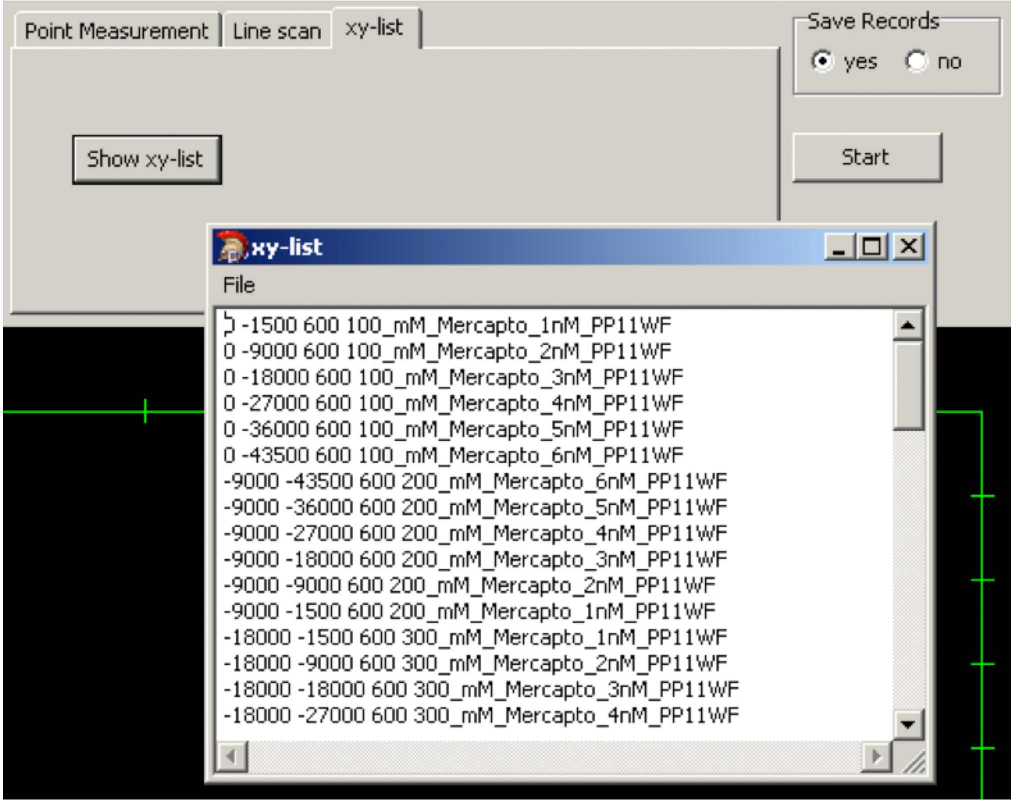


Figure 56: Assembled stage, holder and wellplate on confocal microscope for automated single-molecule measurements. Red arrows indicate the adjustment screws to align the wellplate on the microscope. To avoid contamination and evaporation, the wells can be sealed with a transparent, self sealing film. The excitation laser can be seen as a blue light in some wells in the middle of the plate.

The automated stage is controlled by the remote controller shown in figure 57. This device can either be controlled by a joystick or remotely by software on the data acquisition computer. For the alignment of the wellplate and to get the coordinates needed to write the automated scanning script, the information displayed on the LCD screen can be used. Always check that the zero point is set correctly. Incorrect parameters can lead to a collision between the metal stage and the objective which may have fatal consequences for the objective lens.

Figure 57: Remote controller for the automatic stage. The display shows the current position of the stage. It can be used for alignment and for programming the automated measurement job. The "CLEAR" and the "RESET" button can be used to set the axes to zero. This should be done if positioned in the first well. The "SPEED" knob adjusts the repositioning speed by using the manual (JOY-STICK) control. To control the stage remotely using the computer, the "JOY-STICK" button has to be set to "AUTO".

The stage and the corresponding measurements can be controlled via a software made by Daniel Nettels called “Countikon”. In a special configuration window called “xy-list” the coordinates, the length of the measurement and the filename can be entered. Since this application can open .txt files, these job files can also be generated automatically by using an external application such as Wolfram Mathematica. When programming a job, it is important to consider the fact that water immersion objectives are used. Since evaporation and loss of water due to relocation may result in problems with focussing, traveling distances of the stage should be as short as possible. To this end, a meander scheme is most efficient.



```
0 -1500 600 100_mM_Mercapto_1nM_PP11WF
0 -9000 600 100_mM_Mercapto_2nM_PP11WF
0 -18000 600 100_mM_Mercapto_3nM_PP11WF
0 -27000 600 100_mM_Mercapto_4nM_PP11WF
0 -36000 600 100_mM_Mercapto_5nM_PP11WF
0 -43500 600 100_mM_Mercapto_6nM_PP11WF
-9000 -43500 600 200_mM_Mercapto_6nM_PP11WF
-9000 -36000 600 200_mM_Mercapto_5nM_PP11WF
-9000 -27000 600 200_mM_Mercapto_4nM_PP11WF
-9000 -18000 600 200_mM_Mercapto_3nM_PP11WF
-9000 -9000 600 200_mM_Mercapto_2nM_PP11WF
-9000 -1500 600 200_mM_Mercapto_1nM_PP11WF
-18000 -1500 600 300_mM_Mercapto_1nM_PP11WF
-18000 -9000 600 300_mM_Mercapto_2nM_PP11WF
-18000 -18000 600 300_mM_Mercapto_3nM_PP11WF
-18000 -27000 600 300_mM_Mercapto_4nM_PP11WF
```

Figure 58: Screenshot of the software implementation of the stage remote control. An “xy-list” such as the one shown here is a simple .txt file with the control instructions. It has the following file-format: x-position / y-position / measuring time in seconds / filename. Values are separated by spaces. The correct coordinates can be taken from the controller shown in figure 57.

By performing measurements using this automated approach presented here, there are a few important points to keep in mind:

- The displacement between the objective and the wellplate (z-axys) can not be remote controlled in this setup but has to stay fixed during the whole automatic measurement. Therefore, due to the non-planarity of the wellplates, which can be up to 75 µm over the whole plate, closer working distances between objective and the glass bottom of the well-plate are required compared to those used for standard single-molecule experiments. This increases the risk of a direct contact between the objective and the glass of the wellplate, which can lead to scratches on the objective lens. Additionally, dust free working conditions are absolutely required since small particles between the bottom of the glass plate and the objective lens may lead to scratches due to the very close working distance.
- Proper perpendicular alignment of the wellplate relative to the optical axis is crucial for

reliable measurements. Therefore the adjustment using the adjustment screws should be performed very carefully. Usually it is best to measure the height of the bottom glass using the micrometer scale of the focus wheel of the microscope and the back reflection of the laser over several points in the plate.

- Since these wellplates are made of plastic, chemical stability in terms of substances and incubation time is very limited. As an example, the manufacturer (Swissci) specifies a stability of 24h for PBS buffer at room temperature. For a 10% DMSO solution in H2O this time is reduced to 1h. Generally the plates are incompatible with alcohols and many other organic solvents.
- Reuse of the wellplates is absolutely not recommended
- Compared to measurements performed in quartz cells, the amount of background bursts for measurements in the wellplates is significantly increased. Cleaning experiments with various tensides were performed but no improvement was found. Although it is possible to perform single-molecule experiments in these plates, application for publication quality single-molecule data collection is not recommended. Interestingly, the amount of background bursts seems to be rather stable over time.

8.4.3 Results

After an extensive search for substances and substance combinations for enhancing the fluorescence signal and photo-protection using the automated setup introduced in the previous section, we propose the following promising combinations:

- The most successful combination was the mixture of mercaptoethanol and cysteamine, which also allowed to study the collapse of unfolded proteins at increased temperatures. Usually a 1:100 v/v dilution of the liquid mercaptoethanol is used which corresponds to a concentration of 143 mM. For the cysteamin, the optimal concentration has to be determined by the user for the specific system of interest. “Optimal” concentrations ranged from 5-30 mM. An extensive systematic concentration screen of relative concentrations of cysteamine and mercaptoethanol revealed an optimal concentration of 150 mM mercaptoethanol and 5 mM cysteamine.
- Another well performing combination, particularly at high irradiation intensities was mercaptoethanol together with paraquat. 0.1-3 mM paraquat with 300 mM mercaptoethanol resulted in significant photo-protection at increased laser powers.
- Instead of paraquat it was also possible to use chloramphenicol at low concentrations. 2-4 mM chloramphenicol together with 300 mM mercaptoethanol also resulted in an improvement of the signal at higher laser powers. However, impurities and/or autofluorescence of chloramphenicol makes the application more difficult. Additionally there was an issue with long term stability.

It should be re-emphasized that the reported strategies work well for the Alexa 488 / Alexa 594 FRET system in PBS buffer under native conditions. The mercaptoethanol/cysteamin system was also successfully applied in different buffers, at different temperatures and in the presence of GdmCl. If different conditions or fluorescent dyes are used, then there is the need to repeat such a screen. It is also very important to mention that long-term stability of the additives and combinations of additives have to be considered. Since these are redox active systems, side reactions such as oxidation and therefore degradation are to be expected. This is an important consideration for long-term experiments and proper storage of the substances.

References

[1] C. E. Aitken, R. A. Marshall, and J. D. Puglisi. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal*, 94(5):1826–1835, 2008.

[2] A. Albini, E. Fasani, and M. Mella. Pet-Reactions of Aromatic Compounds. *Photoinduced Electron Transfer V*, 168:143–173, 1993.

[3] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(96):223–230, 1973.

[4] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences*, 47:1309–1314, 1961.

[5] R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, and W. A. Eaton. Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proceedings of the National Academy of Sciences*, 104(48):18964–18969, 2007.

[6] R. B. Best and J. Mittal. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *The Journal of Physical Chemistry B*, 114(46):14916–14923, 2010.

[7] M. Cai, R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Structural Biology*, 4(7):567–577, 1997.

[8] K. Carayon, H. Leh, E. Henry, F. Simon, J.-F. Mouscadet, and E. Deprez. A cooperative and specific DNA-binding mode of HIV-1 integrase depends on the nature of the metallic cofactor and involves the zinc-containing N-terminal domain. *Nucleic acids research*, 38(11):3692–3708, 2010.

[9] R. A. Cardullo, S. Agrawal, C. Flores, P. C. Zamecnik, and D. E. Wolf. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences*, 85(23):8790–8794, 1988.

[10] H. Chen, S. S. Ahsan, M. B. Santiago-Berrios, H. D. Abruña, and W. W. Webb. Mechanisms of quenching of Alexa fluorophores by natural amino acids. *Journal of the American Chemical Society*, 132(21):7244–7245, 2010.

[11] T. Cordes, J. Vogelsang, and P. Tinnefeld. On the mechanism of Trolox as antiblinking and antibleaching reagent. *Journal of the American Chemical Society*, 131(14):5018–5019, 2009.

[12] M. Dahan, A. Deniz, T. Ha, D. Chemla, P. Schultz, and S. Weiss. Ratiometric measurement and identification of single diffusing molecules. *Chemical Physics*, 247(1):85–106, 1999.

[13] R. Dave, D. S. Terry, J. B. Munro, and S. C. Blanchard. Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophysical Journal*, 96(6):2371–2381, 2009.

[14] A. P. de Silva, H. Q. N. Gunaratne, T. Gunnlaugsson, A. J. M. Huxley, C. P. McCoy, J. T. Rademacher, and T. E. Rice. Signaling Recognition Events with Fluorescent Sensors and Switches. *Chemical Reviews*, 97(5):1515–1566, 1997.

[15] A. A. Deniz, M. Dahan, J. R. Grunwell, T. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss, and P. G. Schultz. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proceedings of the National Academy of Sciences*, 96(7):3670–3675, 1999.

[16] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss. Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proceedings of the National Academy of Sciences*, 97(10):5179–5184, 2000.

[17] A. A. Deniz, T. A. Laurence, M. Dahan, D. S. Chemla, P. G. Schultz, and S. Weiss. Ratiometric single-molecule studies of freely diffusing biomolecules. *Annual Review of Physical Chemistry*, 52:233–253, 2001.

[18] K. Dill and H. Chan. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1):10–19, 1997.

[19] P. S. Dittrich and P. Schwille. Photobleaching and stabilization of fluorophores used for single-molecule analysis with one- and two-photon excitation. *Applied Physics B: Lasers and Optics*, 73(8):829–837, 2001.

[20] S. Doose, H. Neuweiler, and M. Sauer. A close look at fluorescence quenching of organic dyes by tryptophan. *Chemphyschem*, 6(11):2277–2285, 2005.

[21] C. Eggeling, S. Berger, L. Brand, J. Fries, J. Schaffer, A. Volkmer, and C. Seidel. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *Journal of Biotechnology*, 86(3):163–180, 2001.

[22] M. Ehrenberg and R. Rigler. Fluorescence correlation spectroscopy applied to rotational diffusion of macromolecules. *Quarterly Reviews of Biophysics*, 9(1):69–81, 1976.

[23] T. Förster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437(1-2):55–75, 1948.

[24] K. Gast, D. Zirwer, H. Damaschun, U. Hahn, M. Müller-Frohne, M. Wirth, and G. Damaschun. Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: a dynamic light scattering study. *FEBS Letters*, 403(3):245–248, 1997.

[25] C. Gell, D. Brockwell, and D. A. Smith. *Handbook of single molecule fluorescence spectroscopy.* Oxford University Press, USA, 2006.

[26] A. Gershenson and L. M. Gierasch. Protein folding in the cell: challenges and progress. *Current Opinion in Structural Biology*, 21(1):32–41, 2011.

[27] I. Gopich and A. Szabo. Theory of photon statistics in single-molecule Förster resonance energy transfer. *The Journal of Chemical Physics*, 122(1):14707, 2005.

[28] I. V. Gopich, D. Nettels, B. Schuler, and A. Szabo. Protein dynamics from single-molecule fluorescence intensity correlation functions. *The Journal of Chemical Physics*, 131(9):095102, 2009.

[29] I. V. Gopich and A. Szabo. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *The Journal of Chemical Physics*, 124(15):154712, 2006.

[30] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, 93(13):6264–6268, 1996.

[31] T. Ha, I. Rasnik, W. Cheng, H. P. Babcock, G. H. Gauss, T. M. Lohman, and S. Chu. Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, 419(6907):638–641, 2002.

[32] D. Haenni. Development and Application of Single-Molecule Total Internal Reflection Microscopy for the Quantitative Analysis of Peptide Aggregation. Master's thesis, University of Zurich, Zurich, 2006.

[33] G. Haran. Single-molecule fluorescence spectroscopy of biomolecular folding. *Journal of Physics-Condensed Matter*, 15(32):R1291–R1317, 2003.

[34] F. U. Hartl. Molecular chaperones in cellular protein folding. *Nature*, 381(6583):571–580, 1996.

[35] F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, and B. Schuler. Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angewandte Chemie (International ed. in English)*, 47(33):6184–6188, 2008.

[36] F. Hillger, D. Nettels, S. Dorsch, and B. Schuler. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *Journal of Fluorescence*, 17(6):759–765, 2007.

[37] P. Hindmarsh and J. Leis. Retroviral DNA integration. *Microbiology and molecular biology reviews : MMBR*, 63(4):836–43– table of contents, 1999.

[38] M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, and H. Grubmüller. Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PloS one*, 6(5):e19791, 2011.

[39] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*, 104(1):105–110, 2007.

[40] A. Hoffmann, D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, and B. Schuler. Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: recurrence analysis of single particles (RASP). *Physical Chemistry Chemical Physics*, 13(5):1857–1871, 2011.

[41] H. Hofmann, F. Hillger, S. H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. A. Lipman, and B. Schuler. Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proceedings of the National Academy of Sciences*, 107(26):11793–11798, 2010.

[42] A. L. Horwich, A. C. Apetri, and W. A. Fenton. The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS Letters*, 583(16):2654–2662, 2009.

[43] A. L. Horwich, S. G. Burston, H. S. Rye, J. S. Weissman, and W. A. Fenton. Construction of single-ring and two-ring hybrid versions of bacterial chaperonin GroEL. *Methods in Enzymology*, 290:141–146, 1998.

[44] A. L. Horwich, W. A. Fenton, E. Chapman, and G. W. Farr. Two Families of Chaperonin: Physiology and Mechanism. *Annual Review of Cell and Developmental Biology*, 23(1):115–145, 2007.

[45] W. A. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, and F. U. Hartl. Identification of in vivo substrates of the chaperonin GroEL. *Nature*, 402(6758):147–154, 1999.

[46] A. Ikai and C. Tanford. Kinetic Evidence for Incorrectly Folded Intermediate States in the Refolding of Denatured Proteins. *Nature*, 230(5289):100–102, 1971.

[47] J. Jacob, R. S. Dothager, P. Thiyagarajan, and T. R. Sosnick. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *Journal of Molecular Biology*, 367(3):609–615, 2007.

[48] J. Jacob, B. Krantz, R. S. Dothager, P. Thiyagarajan, and T. R. Sosnick. Early collapse is not an obligate step in protein folding. *Journal of Molecular Biology*, 338(2):369–382, 2004.

[49] Y. Jia, D. Talaga, W. Lau, H. Lu, and W. DeGrado. Folding dynamics of single GCN-4 peptides by fluorescence resonant energy transfer confocal microscopy. *Chemical Physics*, 247:69–83, 1999.

[50] P. Kask, P. Piksarv, U. Mets, M. Pooga, and E. Lippmaa. Fluorescence correlation spectroscopy in the nanosecond time range: rotational diffusion of bovine carbonic anhydrase B. *European Biophysics Journal*, 14(4):257–261, 1987.

[51] P. Kask, P. Piksarv, M. Pooga, U. Mets, and E. Lippmaa. Separation of the rotational contribution in fluorescence correlation experiments. *Biophysical Journal*, 55(2):213–220, 1989.

[52] H. Katou, M. Hoshino, H. Kamikubo, C. A. Batt, and Y. Goto. Native-like beta-hairpin retained in the cold-denatured state of bovine beta-lactoglobulin. *Journal of Molecular Biology*, 310(2):471–484, 2001.

[53] S. Kawato, K. Kinosita, and A. Ikegami. Dynamic structure of lipid bilayers studied by nanosecond fluorescence techniques. *Biochemistry*, 16(11):2319–2324, 1977.

[54] M. J. Kerner, D. J. Naylor, Y. Ishihama, T. Maier, H.-C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, and F. U. Hartl. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell*, 122(2):209–220, 2005.

[55] Y. Kim, S. O. Ho, N. R. Gassman, Y. Korlann, E. V. Landorf, F. R. Collart, and S. Weiss. Efficient site-specific labeling of proteins via cysteines. *Bioconjugate Chemistry*, 19(3):786–791, 2008.

[56] D. K. Klimov, D. Newfield, and D. Thirumalai. Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proceedings of the National Academy of Sciences*, 99(12):8019–8024, 2002.

[57] O. Krichevsky and G. Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*, 65(2):251–297, 2002.

[58] F. Krieger, B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *Journal of Molecular Biology*, 332(1):265–274, 2003.

[59] E. V. Kuzmenkina, C. D. Heyes, and G. U. Nienhaus. Single-molecule FRET study of denaturant induced unfolding of RNase H. *Journal of Molecular Biology*, 357(1):313–324, 2006.

[60] J. R. Lakowicz. *Principles of fluorescence spectroscopy.* Springer, 2006.

[61] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proceedings of the National Academy of Sciences*, 97(13):7220–7225, 2000.

[62] C. Levinthal. Are there pathways for protein folding? *Journal of Medical Physics*, 65(1):44–45, 1968.

[63] C. Levinthal. *How proteins fold graciously.* Mössbauer Spectroscopy in Biological Systems, Illinois, university of illinois press edition, 1969.

[64] Z. Lin, D. Madan, and H. S. Rye. GroEL stimulates protein folding through forced unfolding. *Nature Structural & Molecular Biology*, 15(3):303–311, 2008.

[65] G. Lipari and A. Szabo. Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophysical Journal*, 30(3):489–506, 1980.

[66] D. Lucent, V. Vishal, and V. S. Pande. Protein folding under confinement: a role for solvent. *Proceedings of the National Academy of Sciences*, 104(25):10430–10434, 2007.

[67] G. I. Makhatadze and P. L. Privalov. Protein interactions with urea and guanidinium chloride. A calorimetric study. *Journal of Molecular Biology*, 226(2):491–505, 1992.

[68] J. Martin, T. Langer, R. Boteva, A. Schramel, A. L. Horwich, and F. U. Hartl. Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. *Nature*, 352(6330):36–42, 1991.

[69] K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, and W. A. Eaton. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of the National Academy of Sciences*, 104(5):1528–1533, 2007.

[70] W. E. Moerner. New directions in single-molecule imaging and analysis. *Proceedings of the National Academy of Sciences*, 104(31):12596–12602, 2007.

[71] W. E. Moerner and D. P. Fromm. Methods of single-molecule fluorescence spectroscopy and microscopy. *Review of Scientific Instruments*, 74(8):3597–3619, 2003.

[72] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, and B. Schuler. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*, 107(33):14609–14614, 2010.

[73] I. Munro, I. Pecht, and L. Stryer. Subnanosecond motions of tryptophan residues in proteins. *Proceedings of the National Academy of Sciences*, 76(1):56–60, 1979.

[74] D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences*, 104(8):2655–2660, 2007.

[75] D. Nettels, A. Hoffmann, and B. Schuler. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *The Journal of Physical Chemistry B*, 112(19):6137–6146, 2008.

[76] D. Nettels, S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, and B. Schuler. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proceedings of the National Academy of Sciences*, 106(49):20740–20745, 2009.

[77] H. Neuweiler, C. M. Johnson, and A. R. Fersht. Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. *Proceedings of the National Academy of Sciences*, 106(44):18569–18574, 2009.

[78] H. Neuweiler, M. Löllmann, S. Doose, and M. Sauer. Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy. *Journal of Molecular Biology*, 365(3):856–869, 2007.

[79] H. Neuweiler, A. Schulz, M. Böhmer, J. Enderlein, and M. Sauer. Measurement of Submicrosecond Intramolecular Contact Formation in Peptides at the Single-Molecule Level. *Journal of the American Chemical Society*, 125(18):5324–5330, 2003.

[80] J. M. Nitsche, H.-C. Chang, P. A. Weber, and B. J. Nicholson. A transient diffusion model yields unitary gap junctional permeabilities from images of cell-to-cell fluorescent dye transfer between Xenopus oocytes. *Biophysical Journal*, 86(4):2058–2077, 2004.

[81] Y. Nomura, T. Masuda, and G. Kawai. Structural analysis of a mutant of the HIV-1 integrase zinc finger domain that forms a single conformation. *Journal of Biochemistry*, 139(4):753–759, 2006.

[82] A. Nöppert, K. Gast, M. Müller-Frohne, and D. Zirwer. Reduced-denatured ribonuclease A is not in a compact state. *FEBS Letters*, 380:179–192, 1996.

[83] E. P. O'Brien, R. I. Dima, B. Brooks, and D. Thirumalai. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *Journal of the American Chemical Society*, 129(23):7346–7353, 2007.

[84] D. Pfiffi, B. A. Bier, C. M. Marian, K. Schaper, and C. A. M. Seidel. Diphenylhexatrienes as photoprotective agents for ultrasensitive fluorescence detection. *The Journal of Physical Chemistry A*, 114(12):4099–4108, 2010.

[85] K. W. Plaxco, I. S. Millett, D. J. Segel, S. Doniach, and D. Baker. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Structural Biology*, 6(6):554–556, 1999.

[86] J. H. Ploegman, G. Drent, K. H. Kalk, W. G. Hol, R. L. Heinrikson, P. Keim, L. Weng, and J. Russell. The covalent and tertiary structure of bovine liver rhodanese. *Nature*, 273(5658):124–129, 1978.

[87] B. E. Poling, J. M. Prausnitz, and J. P. O'Connell. *The Properties of Gases and Liquids*. McGraw-Hill Professional, 5 edition, 2000.

[88] A. Polson. On the diffusion constants of the amino-acids. *Biochemical Journal*, 31(10):1903, 1937.

[89] I. Rasnik, S. A. McKinney, and T. Ha. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods*, 3(11):891–893, 2006.

[90] E. Rhoades, E. Gussakovsky, and G. Haran. Watching proteins fold one molecule at a time. *Proceedings of the National Academy of Sciences*, 100(6):3197–3202, 2003.

[91] P. J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B. M. Wöhrl, T. Restle, R. S. Goody, and C. A. M. Seidel. Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. *Proceedings of the National Academy of Sciences*, 100(4):1655–1660, 2003.

[92] R. Roy, S. Hohng, and T. Ha. A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507–516, 2008.

[93] B. Schuler. Application of single molecule Förster resonance energy transfer to protein folding. *Methods in Molecular Biology*, 350:115–138, 2007.

[94] B. Schuler and W. Eaton. Protein folding studied by single-molecule FRET. *Current Opinion in Structural Biology*, 18(1):16–26, 2008.

[95] B. Schuler, E. A. Lipman, and W. A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002.

[96] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, and W. A. Eaton. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proceedings of the National Academy of Sciences*, 102(8):2754–2759, 2005.

[97] D. J. Selkoe. Folding proteins in fatal ways. *Nature*, 426(6968):900–904, 2003.

[98] E. Sherman. Coil-globule transition in the denatured state of a small protein. *Proceedings of the National Academy of Sciences*, 103(31):11539–11543, 2006.

[99] G. Smith. Kinetics of amine modification of proteins. *Bioconjugate Chemistry*, 17(2):501–506, 2006.

[100] L. Song, C. A. Varma, J. W. Verhoeven, and H. J. Tanke. Influence of the triplet excited state on the photobleaching kinetics of fluorescein in microscopy. *Biophysical Journal*, 70(6):2959–2968, 1996.

[101] A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Müller-Späth, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences*, 2012.

[102] T. R. Sosnick and D. Barrick. The folding of single domain proteins - have we reached a consensus? *Current Opinion in Structural Biology*, 21(1):12–24, 2011.

[103] I. Steinberg. Nonradiative Energy Transfer in Systems in Which Rotatory Brownian Motion Is Frozen. *Journal of Chemical Physics*, 48(6):2411–2413, 1968.

[104] I. Z. Steinberg. Long-range nonradiative transfer of electronic excitation energy in proteins and polypeptides. *Annual Review of Biochemistry*, 40:83–114, 1971.

[105] L. Stryer. Fluorescence Energy Transfer as a Spectroscopic Ruler. *Annual Review of Biochemistry*, 47(1):819–846, 1978.

[106] L. Stryer and R. P. Haugland. Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Sciences*, 58(2):719–726, 1967.

[107] A. Szabo. Theory of Fluorescence Depolarization in Macromolecules and Membranes. *Journal of Chemical Physics*, 81(1):150–167, 1984.

[108] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *The Journal of Chemical Physics*, 72(8):4350, 1980.

[109] F. Takagi, N. Koga, and S. Takada. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proceedings of the National Academy of Sciences*, 100(20):11367–11372, 2003.

[110] D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, and R. M. Hochstrasser. Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proceedings of the National Academy of Sciences*, 97(24):13021–13026, 2000.

[111] S. Tandon and P. Horowitz. Detergent-assisted refolding of guanidinium chloride-denatured rhodanese. The effect of lauryl maltoside. *The Journal of Biological Chemistry*, 261(33):15615–15618, 1986.

[112] Y.-C. Tang, H.-C. Chang, K. Chakraborty, F. U. Hartl, and M. Hayer-Hartl. Essential role of the chaperonin folding compartment in vivo. *Embo Journal*, 27(10):1458–1468, 2008.

[113] T. Tezuka-Kawakami, C. Gell, D. J. Brockwell, S. E. Radford, and D. A. Smith. Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophysical Journal*, 91(5):L42–4, 2006.

[114] D. Thirumalai and G. H. Lorimer. Chaperonin-mediated protein folding. *Annual Review of Biophysics and Biomolecular Structure*, 30:245–269, 2001.

[115] T. Y. Tsong, R. L. Baldwin, and E. L. Elson. The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proceedings of the National Academy of Sciences*, 68(11):2712–2715, 1971.

[116] V. N. Uversky, J. R. Gillespie, and A. L. Fink. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 41(3):415–427, 2000.

[117] R. Vabulas, S. Raychaudhuri, M. Hayer-Hartl, and F. U. Hartl. Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harbor Perspectives in Biology*, 2(12), 2010.

[118] P. V. Viitanen, A. A. Gatenby, and G. H. Lorimer. Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. *Protein Science*, 1(3):363–369, 1992.

[119] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, and P. Tinnefeld. A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angewandte Chemie (International ed. in English)*, 47(29):5465–5469, 2008.

[120] M. Wahl, H.-J. Rahn, T. Röhlicke, G. Kell, D. Nettels, F. Hillger, B. Schuler, and R. Erdmann. Scalable time-correlated photon counting system with multiple independent input channels. *The Review of Scientific Instruments*, 79(12):123113, 2008.

[121] M. R. Wasielewski. Photoinduced electron transfer in supramolecular systems for artificial photosynthesis. *Chemical Reviews*, 92(3):435–461, 1992.

[122] S. Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–1683, 1999.

[123] A. Weller. Electron-transfer and complex formation in the excited state. *Pure and Applied Chemistry*, 16(1):115–124, 1968.

[124] J. Widengren, A. Chmyrov, C. Eggeling, P.-A. Löfdahl, and C. A. M. Seidel. Strategies to improve photostabilities in ultrasensitive fluorescence spectroscopy. *The Journal of Physical Chemistry A*, 111(3):429–440, 2007.

[125] F. Wilkinson, D. J. McGarvey, and A. F. Olea. Excited Triplet State Interactions with Molecular Oxygen: Influence of Charge Transfer on the Bimolecular Quenching Rate Constants and the Yields of Singlet Oxygen [O*2(1.DELTA.g)] for Substituted Naphthalenes in Various Solvents. *The Journal of Physical Chemistry*, 98(14):3762–3769, 1994.

[126] P. E. Wright and H. J. Dyson. Linking folding and binding. *Current Opinion in Structural Biology*, 19(1):31–38, 2009.

[127] F. Zosel. Protein folding in a chaperonin cage. Master's thesis, Zurich, 2010.

Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CD	Circular dichroism
CW	Continuous wave
Cys	Cysteine
DLS	Dynamic light scattering
DSE	Denatured state ensemble
ESI-MS	Electrospray ionization mass spectrometry
FCS	Fluorescence correlation spectroscopy
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
GdmCl	Guanidinium hydrochloride
Gly	Glycine
GroEL/ES	Bacterial chaperone GroEL with co-chaperone GroES
His-Tag	Hexahistidine Tag
HPLC	High-performance liquid chromatography
IDPs	Intrinsically disordered proteins
IN	N-terminal domain of HIV-1 integrase
IRF	Instrumental response function
MD	Molecular Dynamics
ns-FCS	High resolution fluorescence correlation spectroscopy
NSOM	Near-field scanning optical microscopy
PET	Photoinduced electron transfer
RASP	Recurrence analysis of single particles
SAXS	Small angle X-ray scattering
Ser	Serine
SPAD	Single photon avalanche photodiodes
SR1	Single-ring variant of GroEL
SVD	Singular value decomposition
TFA	Trifluoroacetic acid

Trp	L-Tryptophan
Csp*Tm*	Cold shock protein from Thermotoga maritima
R_g	Radius of gyration

Acknowledgements

During my work I had the opportunity to work with many exciting people. For their inspiring influence, guidance, help and contributions I would like to thank:

- Prof. Ben Schuler for giving me the opportunity to work in his group and for allowing me to work freely on various challenging projects. Additionally I have to thank him for his constant support and for providing a fantastic research environment.
- Prof. Amedeo Caflisch and Prof. Peter Hamm for joining my thesis committee and for having the endurance to read through this work.
- All the former and current members of the Schuler group for providing a very pleasant working environment and for many helpful discussions: Frank Hillger, Armin Hoffmann, Luc Reymond, Daniel Nettels, Franziska Zosel, Hagen Hofmann, Sonja Müller-Späth, Jennifer Clark, Philipp Schütz, Ruth Kellner, Bengt Wunderlich, Iwo König, René Wuttke, Stephan Benke, Frank Küster, Allesandro Borgia, Andrea Soranno, Andrea Holla, Madeleine Borgia and Sandeep Sharma.
- Dr. Daniel Nettels for innumerable hours of discussions and generous support. Especially I would like to thank him for providing the analysis tools, building and maintenance of the instruments and last but not least for revising my thesis.
- Dr. Sonja Müller-Späth and Dr. Armin Hoffman for initially sharing their knowledge about single-molecule spectroscopy and for introducing me to the group.
- Dr. Luc Reymond for all his advices on organic chemistry and very helpful support.
- Dr. Frank Hillger and Dr. Hagen Hofmann for all the exciting experiments on the GroEL/ES system.
- Franziska Zosel for her master's project on the HIV-1 integrase and her continuous support on the quenching project.
- Dr. Jennifer Clark for proofreading and commenting on my thesis.
- Brigitte Buchli for performing some of the photo protection screens during a semester project and therefore contributing to the improvement of our single-molecule experiments.
- The department of Biochemistry of the University of Zurich for providing a fantastic infrastructure. Especially I like to thank Steve Rast and Stefan Klauser from the IT-service for their fast and reliably service.

Most importantly I would like to thank all my friends and especially my family for their ongoing and very encouraging support. I also like to thank Sandra Lehmann for her valuable support and for sharing her time with me.

Curriculum Vitae

Personal Information

Dominik HÄNNI
born 5th June 1981 in St.Gallen, Switzerland
Place of origin: Toffen, Bern
Swiss citizen

Education

December 2006 - Present	PhD student at the Department of Biochemistry, University of Zurich
November 2006	Master of Science in Biochemistry, Chemistry/Biophysics Track at the University of Zurich, Minor subject: Bioinformatics, Title of master thesis: „Development and Application of Single-Molecule Total Internal Reflection Fluorescence Microscopy for the Quantitative Analysis of Peptide Aggregation“
October 2005 - November 2006	Master studies in Biochemistry and Bioinformatics at the University of Zurich
October 2005	Bachelor of Science in Biochemistry, Chemistry/Biophysics track at the University of Zurich
October 2002 - October 2005	Bachelor studies in Chemistry/Biochemistry at the University of Zurich
August 1997 - June 2001	Gymnasium Friedberg, Gossau, grammar school specialising in economics